FIELD EXPERIMENTATION WITH FRUIT TREES AND OTHER PERENNIAL PLANTS

by
S. C. Pearce

Technical Communication No. 23 (revised)
of the Commonwealth Bureau of Horticulture
and Plantation Crops

C·A·B

COMMONWEALTH AGRICULTURAL BUREAUX

A good field experiment has to be satisfactory in several ways. The farm superintendent has to be able to manage it on the ground, the statistician must be able to carry out a valid analysis of the data and, above all, the experimenter has to obtain precise results relevant to what he wants to know. In this book, first published in 1953, and now fully revised, the author attempts to bring the threads together. Other texts deal more fully with particular aspects; the aim here is integration.

For 37 years the author was a member of the Statistics Section of the East Malling Research Station and saw field experimentation on temperate fruit trees at close quarters, but he became ever more interested in similar problems with tropical crops, so the reference in the title to "other perennial plants" is not perfunctory. The book deals with the place of the field experiment in agricultural research and sets out the usual experimental designs, but with the emphasis that there are more possibilities than most people suspect. There is a chapter on the control of experimental error by the use of covariance and another on plot technique, while the various sorts of experiment, e.g. on fertilizers, varieties, are considered in turn. Special attention is given to problems of data analysis and the general principles of recording and measuring.

£10.00
~~£12.00 ($28.80)~~

CAB INTERNATIONAL
00090442

CABI BIOSCIENCE
EGHAM LIBRARY

FIELD EXPERIMENTATION WITH FRUIT TREES AND OTHER PERENNIAL PLANTS

This and other publications of the
Commonwealth Agricultural Bureaux
can be obtained through any major bookseller
or direct from:

Commonwealth Agricultural Bureaux,
Central Sales,
Farnham Royal,
Slough SL2 3BN,
England

Digitized by the Internet Archive
in 2025

https://archive.org/details/technical-comms-commonwealth-bureau-of-horticulture_1953_23

Aerial photograph of experimental plantations at the East Malling Research Station, showing (1) in the lower left-hand corner, a plum tree-shaping and pruning trial, (2) above on the right, an apple spacing trial planted in fans and (3), below it, an apple interstock trial. Above the fan-trial is (4) another spacing trial.

Jacket illustration and frontispiece reproduced from an aerial photograph supplied by Hunting Surveys Ltd, Boreham Wood, Herts.

FIELD EXPERIMENTATION WITH FRUIT TREES AND OTHER PERENNIAL PLANTS

BY

S. C. PEARCE

East Malling Research Station, Maidstone, Kent, England

Technical Communication No. 23 Second Edition (Revised)
Commonwealth Bureau of Horticulture and Plantation Crops
East Malling, Maidstone, Kent

COMMONWEALTH AGRICULTURAL BUREAUX

First published in 1953 *by the*
Commonwealth Agricultural Bureaux,
Farnham Royal, Slough, SL2 3BN, England
Second Edition (Revised) 1976

© [Commonwealth Agricultural Bureaux, 1976

ISBN 085198 354 5

All rights reserved. No part of this publication may be reproduced, in any form or by any means, electronically. mechanically, by photocopying, recording or otherwise, without the prior permission of the copyright owner,

PREFACE TO THE SECOND EDITION

It is now some 44 years since T. N. Hoblyn's "Field experiments in horticulture" appeared in 1931 and 22 since the publication of the first edition of this book in 1953. The regularity of interval is coincidental but symbolical, because the developments revealed by successive volumes fall into two parts, each important and together complementary. First there was a period where the chief interest lay in *biometry*. Attention was directed to questions like the sources of experimental error, how far it arose from inadequate sampling and recording, and the extent to which the measurements recorded what they purported to. It is true that much biometrical work appeared after 1953 but mostly it was the answering of questions already posed. The real difficulties in writing the first edition came in Chapter 3, which dealt with the more complicated experimental designs. It brought home to me how inadequate was the thinking then current and it led me into my main research activity ever since, which has been in *methodology*, trying to adapt designs to problems, not problems to designs. As with the biometrical aspects in 1953 I feel that we may at last be moving into fulfilment. It would please me to know that in another 22 years the present Chapter 3 was regarded, if remembered at all, as the last expression of an out-dated approach, soon to be superseded by more flexible thinking.

However, biometry and methodology are not to be seen in antithesis. A research team that takes its land and plant material for granted and ceases to bother about the precision of its experiments not only wastes its resources but any efforts to help it by suggesting improved designs will be unavailing because the problems will be unidentified. No less, people who write about a subject called "experimental design" need real problems or they 'darken counsel by words without understanding'. It is not just that biologists and mathematicians can help one another; rather, in matters of experimentation each is ineffective without the other.

The edition represents the culmination of 37 years at the East Malling Research Station, which I have now left to devote myself to the problems of agricultural experimentation in the tropics, where the need is greater and the problems more baffling. I hope the text will be of service to its readers.

September 1975

S. C. P.

University of Kent at
Canterbury, England

CONTENTS

1. Introduction
10. The rise of statistics .. 1
11. The purpose of field trials .. 2
12. The idea of significance .. 4
13. The range of field trials .. 6
14. Characteristics of perennial plants .. 8
15. Choice of experimental plants .. 10
16. The place of statistics .. 11

2. Some Simple Designs
20. Complete randomization and the option of blocking .. 13
21. Randomized blocks .. 15
22. Latin squares .. 17
23. The changing of treatments in a Latin square .. 19
24. Factorial designs .. 22
25. Split-plots .. 24
26. Criss-cross designs .. 26
27. Unrandomized factors .. 27

3. Some More Complicated Designs
30. Non-orthogonal designs in general .. 29
31. Balanced incomplete blocks and other totally balanced designs .. 32
32. Some other kinds of balance .. 35
33. Row-and-column designs .. 37
34. Change-over designs .. 41
35. Using parts of trees as plots .. 42
36. Confounding .. 43
37. Interlaced blocks .. 47
38. Some disadvantages of complex designs .. 48
39. Nonce designs .. 49

4. Calibration
40. The idea of calibration .. 51
41. Covariance and some possible alternatives .. 52
42. Lay-out of plantations for calibrations .. 54
43. Some experiences with calibration .. 57
44. Trials on commercial plantations .. 60
45. The method of adjustment by neighbouring plots .. 60

5. Blocks, Plots and Replications
50. Size and shape of blocks .. 63
51. Size and shape of plots .. 65
52. Review of uniformity trials, etc. .. 68
53. The use of guard rows .. 71
54. Degree of replication .. 73
55. Formulae for determining the number of replicates .. 74

6. Trial Design in Relation to the Subject of Investigation
60. Fertilizer trials .. 76
61. Trials of methods of soil management including irrigation .. 79
62. Spraying and dusting trials .. 80

6. Trial Design in Relation to the Subject of Investigation (*cont.*)
63. Trials of varieties and rootstocks 81
64. Trials of pruning and tree formation 83
65. Trials of commercial systems 84
66. Spacing trials 85
67. Trials of the spread of diseases 91
68. Trials with several kinds of treatment 94

7. The Analysis of Results
70. Some general considerations 95
71. The analysis of a single experiment 97
72. Significant differences and some alternatives 99
73. Heterogeneous errors and transformations 101
74. The analysis of several years' results 105
75. Experiments at several sites 108

8. Some Mishaps and Remedies
80. Errors and blunders 110
81. Subsequent classification and pseudo-variates 111
82. Missing trees and plants 111
83. Missing plots 112
84. Some consequences of fitted values 116
85. Mixed-up plots 117
86. Errors in the application of treatments 117

9. The Measurement of Perennial Plants
90. General survey of the problem 119
91. Location and conditions of recording 121
92. Sampling methods 122
93. Estimates and categories 124
94. Indirect measurements 125
95. Measurements of size and growth 127
96. Measurements of fruit and blossom 129
97. Records of flavour and consumer acceptability 131
98. The importance of good recording 133

Appendix I 134
The use of computers in agronomic research

Appendix II 138
Multivariate methods in agronomic research

Appendix III 147
Calculating the analysis of variance

Appendix IV 155
The analysis of covariance with a note on the partitioning of the treatment variation

References 167

Subject Index 180

LIST OF FIGURES

FIGURE		PAGE
I.	Design with complete randomization	13
II.	Design in which the number of replicates exceeds the number of blocks	14
III.	Designs in randomized blocks. (a) Six blocks of two treatments. (b) Four blocks of six treatments	16
IV.	Example of a 5×5 Latin square	17
V.	Examples of double 4×4 Latin squares. (a) No tying. (b) Tied	18
VI.	Example of a split-plot design	25
VII.	Example of a criss-cross design	27
VIII.	Example of an orthogonal design	29
IX.	Example of a balanced incomplete block design	30
X.	Example of design with two proportionate treatments (A and B)	31
XI.	Use of guard trees for a subsidiary trial	31
XII.	Plan of a spraying trial designed in balanced incomplete blocks	33
XIII.	Example of a Youden square	38
XIV.	Example of a design of type O : OT (i.e. the treatments are disposed orthogonally to rows and in total balance with respect to columns)	38
XV.	Example of a design of type O : OG (i.e. the treatments are disposed orthogonally to rows and in a group-divisible design with respect to columns)	39
XVI.	Example of a design of type O : TT (i.e. the treatments are disposed in total balance with respect to both rows and columns)	40
XVII.	Example of a design of type O : SS (i.e. the treatments are disposed in supplemented balance with respect to both rows and columns)	40
XVIII.	Example of a trial with permanent and temporary trees in interlaced blocks	47
XIX.	Interlaced blocks with partial confounding to permit thinning of a trial with one-tree plots	48
XX.	Example of a herbicide trial designed in supplemented balance	49
XXI.	Examples of calibration areas designed in stripes and tiles	55
XXII.	A system of blocks to eliminate headland effects	64
XXIII.	Alternative methods of forming blocks	64
XXIV.	Examples of heterogeneous headlands	72
XXV.	Different types of internal guard rows. (a) Half-guarding. (b) Single-guarding with and without alley. (c) Double-guarding with and without alley	73
XXVI.	Possible successive stages in the evolution of a design with tentative treatments	78
XXVII.	Examples of different types of systematic spacing trials	87
XXVIII.	Possible relationships between y, the quantity estimated, and x, its estimator	126

CHAPTER 1

INTRODUCTION

> "I heard the Duke of Albemarle's chaplain make a mighty simple sermon: among other things reproaching the imperfection of human learning, he cried: 'All our physicians cannot tell what an ague is, and all our arithmetique is not able to number the days of a man', which, God knows, is not the fault of arithmetique, but that our understandings reach not the thing."
>
> SAMUEL PEPYS, *Diary* (Entry for 5 November 1665)

10. The Rise of Statistics

Over the years research work in agriculture and horticulture has been through many changes, not only in the problems studied but much more in the role of those engaged in it. Once the research worker was a practical farmer, perhaps a member of the landed gentry, with the education to wonder whether the new ideas in chemistry and engineering could help him to grow better crops. Slowly the scientific element in his work increased, but he could still say, to quote some favourite words of Sir Ronald Hatton, "the field is our laboratory". Now the task may be the effective use of an electron microscope or the chemical analysis of spray deposits or the identification of species of bacteria. So long as the motivation comes from the field, it does not matter much where the laboratory is or what it contains, but such changes necessarily lead to different kinds of data and means of dealing with them.

Along with the changes in problems there have been parallel changes in statistical techniques. Indeed, the rise of agricultural research and statistical method were nearly contemporary, both coming from the urge of the early nineteenth century to apply scientific method to the affairs of everyday life. The badge of the Royal Statistical Society, founded in London in 1834, is still a sheaf of wheat—an allegorical one admittedly—symbolizing the garnering of facts, but it was relevant in another way because agriculture was the basis of much prosperity. Discussion of the Corn Laws was fierce and anyone who was proposing to study the wealth of the nation must start with agriculture. Despite their common origin and their evident need of one another, agriculturist and statistician tended to drift apart. For one thing, each became more specialized, the one basing himself on biological science and the other on mathematics, so that they began to find one another incomprehensible; for another their roles began to conflict.

Originally the Royal Statistical Society had a motto as well as a badge. On a band round its middle the wheatsheaf, already reaped, bore the words, *Aliis exterendum* ("Let others do the threshing"), but later it was realized that collecting and presenting data are not enough; it is necessary to extract sense from them as well. Thereafter the statistician appeared to encroach upon the province of other people by appealing to specialized, almost esoteric, considerations with little evident reference to those of practical importance. Indeed, it almost seemed as if earlier roles had been reversed and the experimenter had the uninspiring job of taking measurements on a trial designed by the statistician, who then proceeded to the interpretation of what the data showed. However, that was an illusion, because really their roles are complementary and the best results are obtained when they work in the harmony and mutual appreciation that once were taken for granted. Each has need of the other and it would be foolhardy for them to quarrel.

In this matter horticulturists have been fortunate because one of the first major statistical publications was in their field, namely T. N. Hoblyn's "Field experiments in horticulture" (Hoblyn, 1931) written by one who knew the practical difficulties and who advocated statistical methods while recognizing the importance of non-statistical considerations and appreciating the special needs of different species. For twenty years it was a standard work and the first edition of the present book was intended to perpetuate the practical spirit of that classic text, which in some measure it replaced.

Since that first edition, another twenty years have gone by. If there is anything to be learnt from the period it is the continuing need for co-operation between biologist and statistician. In these sophisticated days it is not enough for a horticulturist to state his impressions without supporting them with data that have been subjected to rigorous examination. Neither should it be enough for a statistician to propound mathematics in a vacuum, although some seem to flourish thereby, because the highly formalized world of the mathematician is not the varied and exciting world of the biologist. It is all very well to say: "Let the yields be normally distributed", but whether they are or not is a matter to be tested, not postulated. Each needs the help of the other. The horticulturist needs the service, while the statistician who applies himself to real problems will be rewarded by concepts as satisfyingly elegant as any that come into the mind unstimulated by natural phenomena.

The biologist who wishes to consult a statistician could well eavesdrop on what they say about him, whether seriously (Finney, 1956a, 1968; Cox, 1968; Sprent, 1970) or facetiously (Anon., 1949). Also, starting with the classic texts of Fisher (1934, 1935) a number of books have been written to explain statistical methods in a biological context (Goulden, 1952; Finney, 1953a, 1955; Snedecor, 1956; Bailey, 1959; Finney, 1962; Pearce, 1965a; Campbell, 1967; Bliss, 1967, 1970) and many have gone into later editions. A number of more general statistical books also are of value (Cochran and Cox, 1950; Quenouille, 1953; Federer, 1955; Cox, 1958) and it is worth studying some of the very early texts (Engledow and Yule, 1926; Tippett, 1931; Wishart and Sanders, 1936) which often make the essential points more forcibly than some more complex, modern publications. Also, there are a number of general papers that can usefully be studied (Yates, 1946; Finney, 1947; Fisher, 1948; Nelder, 1956; Hoblyn, 1958). The summary by Mayer *et al.* (1959) of methods used with grapes also is of interest, as are the comments of Sutton and Potter (1954).

11. The Purpose of Field Trials

The object of an experiment is to obtain data that will elucidate an obscurity or decide a course of action. Consequently a good experimental design can result only from a precise appreciation of the problem to be solved and no refinement of statistical technique can be a substitute for clear-headedness, however valuable it may be in addition. This communication will set out in some detail those techniques that the author has found useful in fruit tree experimentation, but he would emphasize from the start that nothing he can write will diminish by one iota the thoughtfulness and insight needed to conceive and conclude a successful research programme. His aim is to provide tools and not to decide the task.

At this point a distinction should be made between an experiment and a survey, each of which has its own role. In general, a survey studies the effect of a variable in conjunction with all the other variables associated with it. Thus it might appear that farms that fertilize heavily do not suffer much from insect pests. To anyone offering a spray service, or an administrator anxious to locate foci of infestation, the information could well be important, but it would be unwise to conclude that fertilization of itself leads to reduced insect populations. Indeed, the contrary might well be the case, the lusher growth induced by nitrogen, for example, possibly being especially attractive to sucking species. In the survey, however, high fertilization has been studied in conjunction with its usual concomitants. Thus, farms that fertilize well usually spray well also. They often belong to large owners or companies that invest in better sites selected by scientific means and are managed by men technologically trained rather than by countrymen using traditional methods. In an experiment, on the other hand, the various treatments are applied to areas selected at random, which breaks any usual associations with irrelevant variables. If it then appears that different fertilizer programmes lead to different insect populations the relationship must be assumed to be causative.

At one time it was considered sufficient if an experiment evaluated the mean performance of each treatment, but Sir Ronald Fisher (1934, 1935) revolutionized the approach and nowadays

it is required that each trial shall estimate its own error also. To achieve this end the experimental area is divided into "plots" and the treatments are allocated to them at random in order to avoid bias. Thus, if there are twelve plots, it might be decided to apply treatment A to six of them chosen at random and treatment B to the other six. It is now possible to work out the variation between the twelve data in the form of a "sum of squared deviations". Calling the six values from treatment A, $y_{A1}, y_{A2}, \ldots, y_{A6}$ and those from treatment B $y_{B1}, y_{B2}, \ldots, y_{B6}$, the "total sum of squared deviations" is

$$s\text{-}s = (y_{A1}^2 + y_{A2}^2 + \ldots + y_{A6}^2 + y_{B1}^2 + y_{B2}^2 + \ldots + y_{B6}^2) - G^2/12$$

where G is the grand total of the twelve data. If now attention is concentrated on the differences within treatments, which represent the experimental error, the first component is the sum of squared deviations within treatment A, i.e.

$$(y_{A1}^2 + y_{A2}^2 + \ldots + y_{A6}^2) - T_A^2/6$$

where T_A is the sum of data from treatment A, and similarly within treatment B the component is

$$(y_{B1}^2 + y_{B2}^2 + \ldots + y_{B6}^2) - T_B^2/6$$

Taken together these two quantities provide a measure of the "error sum of squared deviations", $s\text{-}s_e$. Finally the differences between the treatment means gives $s\text{-}s_t$, the "treatments sum of squared deviations", i.e.

$$T_A^2/6 + T_B^2/6 - G^2/12$$

It will be seen that the total sum of squared deviations has been "partitioned" into two components, thus:

$$s\text{-}s = s\text{-}s_e + s\text{-}s_t$$

One of them, $s\text{-}s_e$, is due to error alone; the other, $s\text{-}s_t$, to treatments as well as error. If a way could be found of comparing them, a test would be available for the effectiveness of treatments.

A basis can be found by considering the potential for variation in each. In general, if there are n quantities, there are $(n-1)$ independent "degrees of freedom" between them, i.e. once $(n-1)$ differences between them have been determined, all others are fixed and there are no further options. Thus, given three quantities, if (A − B) is known and (B − C) is known, there is no option about (A − C), which is determined by the first two. Hence, amongst twelve data there are eleven degrees of freedom. One of them can be taken as the difference between the two treatment means, five as representing the variation between the six values for treatment A and five similarly for B. The last two groups together provide ten degrees of freedom for the error. If now each sum of squared deviations is divided by its degrees of freedom to give its "mean square", they will all be on a comparable basis. Hence $ms_e = s\text{-}s_e/10$ is the "error mean square" or "error variance". If there were no effect of treatments, $ms_t = s\text{-}s_t/1$ would provide an alternative measure of the same quantity, but if the treatments do differ in their effect it will be inflated, whereas ms_e will be unaffected. In fact, the extent of an effect of treatments is judged by the quantity F, where $F = ms_t/ms_e$. Since both ms_t and ms_e are estimated quantities that depend upon the randomization, it is not to be expected that they will be exactly equal. It is quite possible for F to be less than one, but if it is large there is evidence for an effect of treatments. Tables (e.g. Fisher and Yates, 1938) show that, for one degree of freedom in the numerator and ten in the denominator, even in the absence of treatment effects F will exceed 4·96 by chance on one occasion in 20, 10·04 on one occasion in 100 and 21·04 on one occasion in 1000. Here then is the basis of a genuine test. The idea of "significance" will be examined further in the next Section. The object here is to emphasize that a well-designed experiment can provide an internal measure of its own consistency and reliability.

To achieve this end there are two essentials. One is "replication". There must be repeated observations on which to base the error. The other is "randomization". If the plots of a treatment are not properly dispersed or if they are dispersed in a pattern, the estimate of error will be biased. Thus, if plots of A and B had alternated like the colours on a chess board, the difference in treatment means would no doubt have been well estimated but the consequent reduction in $s\text{-}s_t$ would have led to a corresponding increase in $s\text{-}s_e$, so the experiment would have appeared less precise rather than more and the F-test would have been biased. Randomization is often a nuisance, but it is essential if the experimental error is to be truly estimated.

The "analysis of variance" obtained above is of value in many ways. For one thing it provides an objective measure of the error and so gives an indication of how well or how badly the experiment was conducted, a high value of $s\text{-}s_e$ suggesting that something went wrong. It is more usually expressed as its square-root, known as the "standard error" of the experiment, because that restores the original scale of measurement which was lost when the data were squared. Sometimes the standard error is expressed as a percentage of the general mean of the data, in which form it is called the "coefficient of variation".

Also important is the "standard error of a treatment mean". The standard error defined above relates to the variation of an individual observation. If n of them are combined to form a mean, it will have a variance equal to the error variance divided by n and consequently a standard error equal to

$$\sqrt{\left(\frac{\text{Error variance}}{n}\right)}$$

Similarly if there are two means, one based on m observations and the other on n, the two sets of observations having no member in common, the "standard error of the difference of means" is

$$\sqrt{\left[\text{Error variance}\left(\frac{1}{m}+\frac{1}{n}\right)\right]}$$

These quantities provide a measure of the precision with which the experiment has estimated means and differences between them. Tables (e.g. Fisher and Yates, 1938) of the quantity t enable limits to be set round estimates. Thus, given ten degrees of freedom for error, a quantity is estimated to within $\pm 2{\cdot}228$ times its standard error with only one chance in 20 of lying outside, to within $\pm 3{\cdot}169$ times with only one chance in 100 of being wrong, while if limits are set at $\pm 4{\cdot}587$ times the true value will be outside on only one occasion in 1000. The observant will note that these values, 2·228, 3·169 and 4·587, are respectively the square roots of 4·96, 10·04 and 21·04, the critical values of F for one degree of freedom in the numerator and the same number (10) in the denominator for corresponding significance levels.

12. The Idea of Significance

Fundamental to the methods just described is the idea of significance, which is so often misunderstood that a note may be helpful. Suppose a difference has been reported as being significant at the level, $P = 0{\cdot}01$. It means that the reader has to choose between two incompatible alternatives, namely, he must accept that the difference was due to the treatments or he must assert that the result is a coincidence, the results just happening to fall out like that, the probability of such a coincidence being one in a hundred. In fact, a significance level informs a would-be sceptic of the probability of such a result arising if there were really nothing in the experimenter's claim that he has brought it about by his treatments.

To clear up some possible misconceptions, the statement does not mean that there is a probability of 0·99 that the difference was due to treatments, nor does it mean that the effect exists on 99 occasions out of 100 but not the other one. It calls upon the reader to make a

choice and tells him the low objective probability of one of the alternatives before him as a reason for choosing the other.

The reader's choice will depend upon the nature of the treatments and the preconceptions of his own mind. Suppose that the treated trees had given a crop greater than that of the controls and it is that difference that is significant. If the treatment had been an invocation and ritual dance to the goddess Pomona, a modern would prefer to believe in the coincidence, but he might believe without question if it had been the application of sulphate of ammonia; in the ancient world contrary choices would perhaps have seemed equally obvious. In fact, although the significance level is itself objectively determined, there is a subjective element in its interpretation. Previous knowledge or collateral evidence in favour of a treatment difference will reduce the level needed for convincement, just as contrary evidence will make belief more difficult. It is even possible that the reader believes already and does not require further evidence.

In general, if P is less than 0·05, the evidence in favour of a genuine difference is accounted strong enough to merit publishing. If P is less than 0·01 the effect is regarded as pretty well established, while if P is less than 0·001 the evidence is considered very strong. By a convention that could well be more widely followed, the first of these levels is indicated by one asterisk, the second by two and the third by three. Provided no one becomes a slave to the levels indicated, a useful summary is provided of the weight of evidence.

There is a tendency in some quarters to decry the concept of significance by saying, quite rightly, that reliance should not be placed upon a single trial but only upon a series. With annual species the contention is quite reasonable for it is easy to repeat an experiment in a later season and perhaps obtain different results. The same is true of short-term treatments like spraying trees, but investigations of varieties and cultural methods take an inconveniently long time anyway and a single trial will itself cover a range of seasons. Of course, where a trial is repeated, the series needs to be interpreted as a whole.

In any event it should be emphasized that repeated significance tests defeat their own object. To return to the subject of surveys, the purpose may be to find features associated with high cropping, disease or other characteristic. If enough characteristics are observed at each site, something will almost certainly be found significant. Given a series of data and a column of random numbers there is one chance in twenty of finding a significant correlation between the two ($P = 0{\cdot}05$). Given twenty columns of random numbers it is to be expected that one of them will fit at that significance level; the same is true of twenty irrelevant features. It is always unwise to repeat significance tests without taking care.

Nevertheless, although the concept of significance is a useful one, the calculation of significance levels is not an end in itself. Too often supervisors of research and editors enquire almost automatically if a difference is "significant", sometimes in ludicrous contexts. For example, an experiment has been conducted upon two varieties chosen for their diversity, the intention being to obtain conclusions of some generality. In such circumstances it is silly to report that the two varieties do differ significantly and just as silly to report that they do not, which sounds more like a condemnation of the experimenter's technique than an assertion of scientific importance. Similar considerations can apply in a fertilizer trial. Often no one is questioning that the dressings will have an effect. The purpose of the trial may only be to find out if the expected increase in crop is large enough to justify the additional cost.

However, the requirement that a man who asserts that there is a difference should show it to be significant is not entirely unreasonable. Its fault is rather that of being too stylized, like the ancient requirement that the man who married the princess should first slay the dragon that was devastating the neighbouring province. The test, whether for a loving bridegroom or a future monarch, was perhaps restricted in scope but not entirely irrelevant. So it is with the discoverer of a "difference". He asserts that it exists and that it is sufficiently large to be of importance, so he clearly has some idea of its magnitude. Also, he claims to have been studying such matters, so he should be conversant with the errors to be expected and therefore to be able

to forecast his error variance. All that remains is for him to carry out an experiment large enough to show the difference as "significant" and everyone will believe him. It is quite fair really.

The trouble with significance tests is not so much their use, or even their abuse or misuse, as the tendency to use them thoughtlessly, introducing them even when other statistical approaches would be more helpful. For example, in the case of the two diverse varieties, no one is advantaged by the information that they are not the same, but it could be helpful to have a list of characters in which they differ with a variance to show the precision of each determination. Again, the experimenter's difficulty may not be in deciding if a certain difference exists but in judging how far it measures up to what he expects or requires. Thus, if he believes that a certain application of fertilizer will result in an increase of 20 per cent in crop, which would well justify the cost, and he is faced with experimental results that show only a 10 per cent improvement, which would not, he has to decide whether to abandon the idea or to do more work to improve his estimate of the advantage. In fact, he does not want to know the probability of getting a 10 per cent increase if the real value is *zero* but of getting it when the real value is *20 per cent*. In statistical language the latter probability is known as the "power" of the test and much has been written about it (e.g. Pearson and Hartley, 1951; Tiku, 1967, 1972). Nomograms have been given by Keuls (1960) and Ferguson (1962). There is a useful paper by Federer (1961).

13. The Range of Field Trials

A field trial is not the only way of conducting experiments and for some purposes it is unstaisfactory, chiefly on account of its being restricted to one locality. A survey on the other hand is wider in application but has disadvantages that have already been mentioned. In fact, the distinction is not clear cut and the one technique merges into the other.

At the one extreme is the formal field trial conducted with full rigour at a single centre. Because scientific staff with relevant skill are probably based on the same centre, advanced methods of applying treatments and recording data are available, while there could be more specialized items of farm machinery available than at an out-station. Also, the trial is often passed by knowledgeable people, who will notice features and developments, which would otherwise go unobserved. In one sense then everything is ideal; in another the whole exercise is of questionable value because it takes place at one site, which necessarily cannot typify the whole area over which the crop is grown.

To resolve this problem it is necessary to ask what might have happened if the trial had been conducted, say, on the other side of the hill, where rainfall is higher. If the answer is that yields would have been generally lower and the incidence of fungal diseases generally higher, but in the main the differences between treatment effects would have been much the same, there need clearly be few reservations about generalizing to both sides of the hill the conclusions reached as to which treatment is best. If, however, the answer is that the pattern of treatment response would have been different, the increased rainfall affecting some treatments more than others, then no generalization would be safe. Therefore, it is never wise to apply the *treatment means* themselves outside the area in which they have been observed unless there is reason to believe that the experimental site is strictly representative of a wider area. Usually it is not, but the assumption may be justified in a uniform plain. However, even where the means cannot be applied generally, it is often reasonable to assume that *differences* between them will have wider generality. That is the ordinary argument for the formal experiment, but there are instances where even this more limited assumption cannot be supported. To use statistical language, the cause lies in the likelihood of there being an interaction of sites and treatments.

If it is agreed that conclusions may well be different at different sites, there is no alternative to the carrying out of a series of trials upon which to base recommendations for the region. The undertaking may be a large one, because each trial will have to be big enough to stand on

its own feet without help from others. It is nonetheless essential if sound recommendations are to be made.

At the other extreme it may be reasonable to have a single block of the experiment at each of a number of sites, as recommended by Marticou and Muller (1964). (As will be explained more fully later, a "block" is a unit of the experiment chosen to be uniform within itself.) Where the area under study is genuinely uniform it should not matter how the blocks are dispersed over it, but results from a range of sites are more convincing than those from only one, providing, as they do, more safeguard against misjudgements. In practice a lot depends upon who is going to manage the dispersed blocks. If everything is to be conducted from the research centre, its unfortunate staff may spend all their time in the car rushing from one site to another. If, on the other hand, growers are asked to help, it may be easier to find eight who will each accept one block of the experiment than to find one who will accept all eight blocks and try to manage what is virtually a full-scale experiment. The approach has been explained excellently by Hopp (1954) and discussed also by Monroe and Mason (1955).

Where blocks are dispersed, difficulties can arise if the situation has been misjudged and some lead to different conclusions from the others. In that case it is always possible to exclude some as being different, but that can amount to the derided practice of discarding all data that do not confirm the experimenter's prejudices. The position is different if the sites giving aberrant results have something in common, e.g. they are all on higher land or on heavier soil, but even then care is needed. No one need be surprised if the heavier soils do indeed appear to have different fertilizer requirements, but the differences cannot have seemed important at the beginning or the investigator would not have taken his range of dispersed sites to be so wide. It is a little late to point them out only when the data cause him embarrassment. Here it should be recalled that, as with significance levels in surveys, an investigator can usually find an association by observing enough characteristics, so unlikely conclusions need confirmation. It may indeed be true that the sites indicating aberrant fertilizer requirements all belong to men born on a Thursday, but such a line of enquiry should be either abandoned as absurd or pursued vigorously as a possible revolutionary discovery. An association of itself provides no justification for the exclusion of sites from the investigation.

An intermediate course is to use a range of sites over an area supposed to be uniform but to place two or three blocks at each. The advantage lies in there now being a formal test procedure for the interaction of sites and treatments and hence a means of deciding more objectively which sites can be accounted different. However, such a test will necessarily be less sensitive than that for the mean effect of the treatments under study, so only a large interaction will be detected. Also, if one is found, it is unlikely that there will be enough data from any single site to establish clearly what can be expected in such conditions.

There is one way in which a dispersed experiment verges on a survey. A query may arise about the conditions in which a spray or a fertilizer application will be advantageous and it is proposed to have pairs of plots at a range of sites, one plot in each pair being treated and the other not, to find out. Clearly it is as reasonable to measure a difference of treatment yields as to measure yield itself and much the same considerations apply. Nevertheless, the resemblance to a dispersed experiment is apparent.

A further criticism is sometimes made of the formal field trial at one site, namely, that it often receives cultural treatments that are quite unrealistic and can therefore lead to conclusions with no applications in commercial practice. The force of the criticism depends upon the purpose of the experiment. A physiologist seeking the relationship between leaf area and fruit development will need trees that are growing well and are either unpruned or pruned according to an exact schedule unlikely to be used in practice. On the other hand, if the need is to increase yields, cultural practices must approximate to those used by practical men. The argument is sometimes used in support of experiments on growers' plantations. For a final trial of a recommendation the argument has some force, especially if it is proposed to disperse the blocks to obtain a range of management. For most purposes, however, it suffices to arrange

matters at the research centre so that visiting growers can recognize the practices used as their own.

14. Characteristics of Perennial Plants

It may well be asked in what respect fruit trees and perennial plants in general are so different from other crops as to need special statistical methods. First of all, of course, they live longer and are therefore more susceptible to mishaps, so greater caution is needed in making plans about them. Also, they are in general larger than annuals and are of greater interest as individuals. Thus, the crop of a single cocoa tree can readily be measured and considered, but wheat is usually studied only in aggregates of several hundred plants, if not of several thousands. This concentration on the individual indicates a further source of variation. The single cocoa tree may be large or small from factors that reside in the tree itself, but an equal area of wheat is made up of many plants, some large and some small, whose varying sizes compensate for one another, leaving only environmental factors, such as soil, to be considered. With the perennial, on the other hand, environmental variation is rarely of sole importance.

To take these points one by one, the greater caution needed with perennials militates against plans that will be ruined if some trees die, while it bids the experimenter bear in mind the time when his trial is finished but the trees will remain—ready, if foresight has been exercised, for a further trial. Also, the greater size of plants means that the experimenter cannot divide up his experimental area as he would often like to. If it measures 45 metres by 30 metres and his trees ought to be planted 5 metres apart and he proposes to use one-tree plots, then he must have 54 plots, nine by six. With smaller plants he could have divided up his land with much more freedom.

An important point concerns the sources of error, a subject upon which much has been written. With perennial plants the experimenter should not assume that variation is positional, i.e. environmental, as his colleague working with annuals may often do—a consideration noted by Hatton a long while ago (Hatton, 1931). Some of it may well derive from the plants themselves. The balance between the two calls for some consideration.

In general, in long-term experiments, the environmental differences will build up with time and small differences in growth rate can be important. It is the same in lending money. Over a short term a small change in the rate of compound interest has little effect, but over a long term it can have a large one. So with plants, anything that leads to lasting differences in the rate of growth must as far as possible be controlled if the experiment is to go on for a long time. Hence, care should be taken with the choice of site to secure uniform land and the design should allow for any differences of soil or shelter that cannot be avoided. Also, plants should be rejected for defects like crooked graft unions and leading shoots damaged by insects, which are likely to lead to lasting differences in growth rates. On the other hand, initial differences in size between plants otherwise healthy and complete are likely to disappear and need not be regarded very seriously. To turn to short-term experiments, however, the greatest care is needed to obtain plants initially uniform and it is wise to raise many more than will be used so that selection can be rigorous, but less attention has to be paid to factors that will affect future growth rates and will take time to become apparent. In making such decisions, however, it should be borne in mind that plants grow quickly in favourable conditions and it is the amount of development that matters. Thus, Moore (1968a) found that the tropical species he studied showed a more rapid build up of environmental variation than temperate ones when judged by elapse of time but not when judged by increase in trunk circumference.

Another consideration in deciding the balance between environmental variation and that due to the plants is the importance of genetic factors. Over a long term, seedlings may be expected to become more diverse, though sometimes they are so closely in-bred that the effect is not serious, while clonal material, if of diverse virus status, can be variable. Nevertheless, for experimental purposes the best clonal material is to be preferred.

A further consideration is the general level of growth. The most uniform set of plants is one in which all are dead; all then have exactly the same growth rate, namely zero. The next best is one in which all are growing vigorously with minimum inhibition. In between, variability is usually higher and is at its maximum when plants are struggling to survive. Disease and minimal fertilizer applications can therefore be potent sources of experimental error. If the investigation requires them, they must be permitted, but due account should be taken of their probable effects. Also, as Brian (1966) points out, graft incompatibility leads to variation.

To sum up, perennial plants have their own problems. They are usually larger and need more to be regarded as individuals. Not least, by their longevity they raise questions about the sources of variation, which may well receive differing answers as the trees develop and some sources wax where others wane.

In addition, when they are removed they leave marked residual effects. The papers of Holland and Greenham (1958, 1962, 1966) and of Holland (1967b, c) can usefully be studied.

An experimenter can reasonably ask how he is to discover the relative importance of various sources of error for his crop in his circumstances. The earliest work was that of Fairfield Smith (1938). He pointed out that if variation comes from the plants and if they have been allocated to their positions at random, the variance of mean per tree as determined from a plot of x trees should be that of the individual plants divided by x, i.e.

$$V_x = \frac{V_1}{x}$$

but in fact it more usually follows the law:

$$V_x = \frac{V_1}{x^b}$$

where b takes some value between 0 and 1, i.e. on account of similarities between neighbours larger plots are less variable than might be expected. Koch and Rigney (1951) examined the implications for relative costs using different plot sizes. Hatheway and Williams (1958) advanced methods of calculating b given data.

The argument was later taken further (Pearce, 1955). Suppose that the variation is in fact made up of two parts, one due to the environment as given by Fairfield Smith's law and the other due to the plants themselves and therefore at random, then

$$V_x = \frac{V_1'}{x^b} + \frac{V_1''}{x}$$

If the second term dominates there will be no correlations between the performances of neighbours, but as the first increases in relative importance, correlations will appear. Already Li and Keller (1951) had noted the relevance of the correlation in performance of successive plants down a row. Freeman (1963) took data from a row and worked out correlation coefficients between (i) adjacent plants, (ii) plants separated by only one other, (iii) plants separated by two others and so on. He then showed how they could be used to estimate b and the ratio of V_1' and V_1''. As might be expected, it appeared that environmental effects built up with time. Freeman's method has been developed by Rives (1969), who made a further point, namely that randomness along a row is statistically much the same as randomness in time, a topic to which much thought has been given. In particular he proposed to use the method of von Neumann (1941), who worked out two measures of variation. For the first he used the ordinary variance, the calculation of which pays no attention to order. He then derived the variance from the differences between successive values. In the random case the resulting value should be twice that obtained by the ordinary method, but if there is association between neighbours it will give less than that. Rives himself applied these various approaches to weights of prunings of grape vines and reached the surprising conclusion that successive plants

along a row were more diverse, not more similar, than if they had been at random. That was perhaps a result of competition, which certainly needs to be considered.

A different approach (Pearce, 1960a) concentrated on relationships in time. Thus, if trees grow without reference to their size at planting, adjustment of growth by initial size will have no effect. If, on the other hand, initial size completely determines subsequent growth, the standard error of size will be proportionately the same at all times and so on. The method was to take the logarithm of trunk circumference as the standard record and to work out:

(i) the standard errors of its current values over a period;
(ii) the standard errors of the difference between current and initial values;
(iii) the standard error in (i) after adjustment by the initial values.

For apples the three graphs exhibited distinctive and reproducible shapes that could be explained by postulating three sources of variation, namely:

(a) the initial variation brought over from the nursery, which had disappeared by the time the trunk girths had attained three or four times their initial values;
(b) a source that built up after planting and rapidly attained its maximum value; that was ascribed to differential planting shock;
(c) a source that built up relentlessly as growth proceeded and ultimately came to dominate all else; that was supposedly the effect of environmental differences.

Further work was done on apples (Pearce and Moore, 1962) and on oranges, peaches, cocoa, cherries and pears (Moore, 1968a). With this wider survey it became apparent that variation did not necessarily build up over the years, though correlations with initial size always disappeared.

15. Choice of Experimental Plants

Quite apart from the need to select within fine or coarse limits, a decision has to be made about the level at which rejections are to be made. Sometimes the experimenter wants good trees and sometimes representative ones.

Where the treatments are inherent in the plants themselves, as with trials of varieties, rootstocks and forms of graft, the trees need to be representative of their kind and ideally selection should be completely at random, excluding only those trees that on account of malformation or palpable defect cannot be expected to grow properly. If uniformity is called for, each kind of tree must set its own limits. Thus, the experimenter might find the mean initial growth of the available plants of each kind, by either sampling or measuring them all, and decide to use only those within 10 per cent of it. If he neglects this precaution and sets the same limits throughout, he may prejudice his results by choosing only large trees of dwarf varieties and undersized trees of vigorous ones.

Where the treatments are applied to trees initially the same, as in a trial of pruning methods or fertilizers, there is no objection to a single set of limits, wide or close as circumstances require, and selecting within them, later allocating the plants to treatments by some random process.

Strict randomness in allocating trees to their field positions is sometimes difficult, though the problems are reduced if plants in the nursery can be given labels of a distinctive colour to show that they have been selected and to indicate the location of each in the field to which they are going. With a large planting, organization can get out of hand, but there is nothing wrong in planting by blocks. The nursery is divided into sections, one for each block of the field trial, and plants are selected for each block from the corresponding section, one block being finished before another is begun. (Indeed, the method has the advantage of allowing for any permanent effects of variable nursery conditions, supposing such to exist.) Whatever departures from randomness are tolerated, there must be no suggestion of assigning different grades of plants to different treatments. The attitude that a tree is only just good enough for the experiment but will be all right as a control is quite reprehensible.

It is natural for trees rejected for experimental use to be allocated to guard rows, but it is important that they should be adequate to do what is required, e.g. intercept spray drift or set up a barrier of roots between two fertilizer plots. Also, a useful role of guards is the equalization of competition among the experimental trees. None of these functions will be served if guard rows are made up of plants rejected on account of their size, so enough material must be raised in the first place to meet all needs.

There are circumstances in which it matters little how the plants are selected, though it is still wise to err on the side of caution. For some species all plants botanically complete end up much the same. Rogers and Edgar (1938) found that to be so for strawberries. Upshall (1939) found that initially larger peach trees grew less well and several investigators have reached the same opinion about apples (Brase and Tukey, 1936; Shaw, 1942; Dorsey and Hough, 1943; Pearce, 1949). With citrus, on the other hand, Webber (1932) found that initial size differences persisted for some years and Lutz (1939) found the same for pecans. Bradford (1944) pointed out that some characteristics of future importance, e.g. crotch angle in apples, are introduced early. Full studies have been made by Cannon (1961) and Neumann (1955). Freeman (1965) similarly studied the effect of different ways of raising cocoa seedlings.

16. The Place of Statistics

In deciding upon a design it should be remembered that statistical considerations, though very important, are never paramount. Whatever design is used, it must be thoroughly practicable, for it is no use producing a scheme on paper that will give endless trouble in the field. If trees are going to need 8 metres to grow in, it is no use planting them 7 metres apart to allow of sufficient replication. Pollinators and shade trees are often a nuisance, but if they are needed they must be provided and in sufficient number to be effective. Statistical requirements are sometimes given such weight that results are based on abnormal trees growing in abnormal conditions and that is a bad mistake.

Neither may statistical considerations be allowed to modify the treatments. If clean cultivation on the farm means ploughing, the experiment must not test hand digging. Again, a knapsack spray on a small tree might have a different effect from commercial spraying in an established orchard, while a broadcast application of fertilizer is not the same as drilling it in. Of course, some trials of a fundamental nature are better conducted using one leaf in a Petri dish than a whole plantation, but in trials that purport to test farm operations there should be no substitute unacceptable to practical men. The excellence of Dyke's (1974) book on the practicalities of experimentation on annuals points the need for a similar text on perennials.

Also it should be borne in mind that mathematical concepts are introduced only as a means of defining and investigating horticultural questions and never for their own sake. It is, of course, true that mathematical ideas can rarely be adequately expressed in words, but where a question has been translated from the language of the botanist into the language of the mathematician, the answer must be of a sort that can be translated back again. It is no good coming to the conclusion that the cubic effect of this interacts with the parabolic effect of something else, the data being in square-root transformation, and supposing that the answer will without further explanation help an investigator who wants to know how much superphosphate he should give when bushes are planted 2000 to the hectare.

Along with this, the trial must be statistically valid and reasonably sensitive to treatment effects. This requirement can no more be relaxed than the others and it is sometimes a severe test of the designer's skill to reconcile them all. In particular, care is needed in avoiding the attitude that a good statistician can analyse anything and consequently almost any design will prove valid if only good enough advice can be obtained eventually. Such confidence can easily be misplaced, especially if the randomization has been restricted to meet practical ends. Some of the more complicated designs were evolved in the first place expressly to make the reconciliation of statistical and practical considerations possible and those can be very useful, but it is better to avoid complexity whenever possible. Above all, it is foolish to indulge in the sort of

statistical virtuosity that supposes any complex design to be better than any simple one and the laying out of a trial to be an exhibition of knowledge. The true aim of the designer should be to obtain the desired information as easily as possible, so complexity of analysis should really go into the debit side of the account, though even so it can sometimes be well worth while.

To sum up, a design must be satisfactory in several respects. Before any trial is proceeded with it must be possible to answer "Yes" to three questions: Is it practicable? Is it statistically sound? Are the treatments really what they purport to be? A failure in any of these respects can be disastrous.

CHAPTER 2

SOME SIMPLE DESIGNS

"And as in races it is not the large stride or high lift that makes the speed; so in business, the keeping close to the matter, and not taking of it too much at once, procureth dispatch."

FRANCIS BACON, *Essay XXV*, "Of Dispatch".

20. Complete randomization and the option of blocking

The simplest of all experimental designs is one in which no attempt is made to allow for position, the treatments being allocated to plots completely at random as was done in the example in Section 11, a possible plan being that in Figure I. It may be objected that such a complete disregard of possible environmental differences is unwise. What would happen if all plots of treatment A were assigned to one end of the area and all plots of B to the other? It is in fact unlikely that such a randomization should arise, though it might. It is, however, true that the lack of any attempt to control environmental variation by dispersal of the treatments is likely to lead to an error variance larger than necessary.

B	A	A	B
A	B	A	A
B	B	B	A

FIGURE I. Design with complete randomization.

It is usual practice to divide the land into "blocks" such that each is environmentally as uniform as possible within itself and to make comparisons only within them. Thus, as in Figure IIIa, it would have been possible to have formed six pairs, each of which received A on one plot and B on the other. Then the only environmental variation to contribute to error would have been that within pairs, not within the area as a whole; consequently the error should be reduced.

To argue thus is quite reasonable but it is necessary to consider also the number of degrees of freedom for error. With the completely randomized design it was ten. If five are to be extracted as belonging to differences between the six pairs (i.e. blocks), there will only be five left for error. It will then be necessary for F to exceed 6·61 to be accounted significant at $P = 0{\cdot}05$, whereas 4·96 would have been sufficient before. In general, in a small trial complete randomization is often better. In such instances the increase in error variance can hardly be large because the total area of the trial is small anyway, while the gain in error degrees of freedom can be of especial value when plots are few. Mostly an experimenter should hope for at least twelve degrees of freedom for error, though more are desirable and numbers like six, five or even four are sometimes unavoidable. When error degrees of freedom are few some care is needed not to squander them on a blocking system that may prove ineffective.

Among other advantages of complete randomization may be mentioned the ease with which it is possible to measure the variation between plots having a particular treatment, a matter that is often of importance as Section 73 will show. When there are no blocks to complicate the issue it is necessary only to work out the mean square of the plot-to-plot variation for each treatment separately to see if they are all about the same. Also, the analysis of results is not made much more difficult by some treatments being represented more frequently than others.

For those who are not prepared to risk all plots of one treatment coming together, but who see the advantage of conserving error degrees of freedom, a good compromise is to use blocks fewer in number than the replicates, as has been done in Figure II. Here the experimental area has been split into two blocks each containing three plots of A chosen at random and three of B. The value of F needed for significance ($P = 0{\cdot}05$) is now 5·12, there being one degree of freedom for blocks, one for treatments and nine for error, so the loss of sensitivity as compared with complete randomization is not serious.

A	A	B	A
B	B	B	A
A	B	A	B

FIGURE II. Design in which the number of replicates exceeds the number of blocks.

Now that the idea of blocks has been introduced, it is important to clarify their role. Primarily they serve two purposes. One is to provide local control of the environment. Some parts of the experimental area will be different from others and it is important that the design should take such variation into account. The differences between blocks are eliminated in the process of statistical analysis, so the estimation of treatment differences and the determination of error depend only on variability within the blocks. It is not essential that all blocks be made up in the same way with respect to treatments, though it saves trouble if they are. Even if they are not, it is possible to allow for differences between them, though it then becomes necessary to apply adjustments to the treatment means, as will be explained in Chapter 3.

The other purpose is to divide the experiment into areas of reasonable size for uniform management. Perhaps a crop is being harvested and it becomes clear as evening approaches that work will have to go into a further day. If the current block is finished and no work started on the next, there will be fewer serious consequences if there is wind during the night that blows off a lot of fruit or rain that causes delay in resumption, because differences between blocks will be allowed for. Similarly, if several people have to record an experiment and it is thought that there could well be subjective differences between them, e.g. in the stage of blossom that each would regard as "fully open", it is better if each can be assigned a complete set of blocks so that their differences in standards shall be eliminated with the other block differences. Strictly speaking it is not necessary to discard the inter-block information. If some blocks contain different treatments from others a comparison of the two sorts of block should provide additional information about treatment effects and ways exist of "recovering" it, nor

should such methods be despised (Yates, 1937, 1938). However, if the inter-block differences are used in the way just recommended, i.e. as a kind of rubbish dump for unwanted sources of variation, it is better not to attempt recovery. In general, recognizable sources of error should be dealt with in one of two ways. Either they should be associated with blocks and so eliminated or, if they must remain intra-block, randomized on to plots so that they do not lead to bias in the estimation of error or the application of significance tests, etc. (It may then be possible to remove them by a covariance adjustment, as described in Appendix IV.) With long-lived species there is much need of a rubbish dump because there are so many sources of variation unforeseeable at the start. Hence, recovery is rarely a useful operation. Also, when, as usually happens, all blocks are made up in the same way, there is no inter-block information concerning treatments to be recovered.

One result of introducing blocks is to cast doubt upon the absolute values of the treatment means. In an agricultural context they should always be regarded with some reserve anyway, because it is difficult to believe that they would have been the same if the experiment had been conducted in another field or planted on a different day or left in the charge of some other person, though it is usually reasonable to assume that differences between treatments would have been much the same. If, however, blocks have been introduced, there will have been restrictions in the choice of material and land and in the deployment of labour to secure uniformity within the blocks, so it is especially hard to believe in the means themselves, though "contrasts" between them may remain plausible.

It may help to define the word "contrasts" more formally. Let an asterisk before the designation of a treatment indicate the mean value to which it gives rise. This mean may have been subjected to a number of adjustments but that does not matter. Then $*A - *B$ represents the contrast between treatments A and B and may be written $(1, -1)$. If the interest should lie in the difference between treatment A on the one hand and treatments B and C regarded as a group on the other, the contrast may be written

$$*A - \tfrac{1}{2}(*B + *C) = *A - \tfrac{1}{2}*B - \tfrac{1}{2}*C$$

or $(1, -\frac{1}{2}, -\frac{1}{2})$. So long as the coefficients sum to zero, the quantity is a "contrast". In a block design only contrasts can be estimated from the intra-block information.

21. Randomized Blocks

To design an experiment in randomized blocks, the total area is divided into as many blocks as there are to be replicates and each block is divided into as many plots as there are to be treatments. Within each block the treatments are assigned to the plots at random, each treatment occurring once and only once in each block, as in Figure III. This is a long-established method. It is not essential for the blocks to adjoin one another. The advantages have been discussed by Rigney (1946).

With perennial species randomized blocks are quite the most usual design and it will be profitable to dwell upon the advantages, for they illustrate very well some fundamental points of trial design with long-lived plants.

First of all, it is an "orthogonal" design. The plots may be classified in two ways, by blocks and by treatments. By orthogonality is meant that, if the plots are grouped according to their blocks, each group will be made up in the same way in respect of treatments, each occurring once in Figure IIIa, but each occurring three times in Figure II. Equally, if they are grouped according to treatments each group will be made up in the same way in respect of representation of blocks. In fact, the trials shown in Figures II and III may both be described as based on "orthogonal two-way classifications". In general, all computing is very much simplified.

The next thing to notice is that the design is "robust". In the course of years, accidents are bound to happen to a trial in a way that is unusual with one that lasts only a short time.

(a)

A	B	B	A
A	B	A	B
B	A	A	B

(b)

C	D	F	E	B	D
A	E	B	F	A	C
A	F	B	A	C	E
E	C	D	D	B	F

FIGURE III. Designs in randomized blocks. (a) Six blocks of two treatments. (b) Four blocks of six treatments.

Thus, some of the treatments may lead eventually to death or abnormality of the trees, while disease may cause havoc in one or more of the blocks and whole plots may be lost. Even if all the trees remain, some of the treatments may prove disappointing so that the experimenter is no longer interested in them. In such circumstances, the robustness of randomized blocks is very useful, for whole blocks and whole treatments may be omitted from the analysis of results without disturbing the basic design, which remains an orthogonal two-way classification. Also, the effect of missing plots is by no means serious (see Sections 83 and 84). Of course, the loss of data must diminish the sensitivity of the trial, but a wise experimenter draws up his design in the first place with something in hand knowing that losses are possible.

A third good property is "flexibility", that is to say, the potentiality of being available for a further set of treatments when the original set is no longer of interest. A fuller discussion of the subject will be left until Section 32, but it may be remarked here that randomized blocks are perhaps the most flexible of all designs.

Against these advantages it can rightly be urged that a trial in randomized blocks, when treatments are numerous, gives rise to a large experimental error on account of the environmental variation being uncontrolled within the large area of a block. For this reason many designs have been suggested in which each block shall contain only some of the treatments and so be smaller. Some of these will be discussed later in Sections 31 and 35; but if much of the variation is inherent in the trees themselves, too much attention can easily be given to the

control of environmental variation. It is not always of paramount importance. Nevertheless, large blocks are to be avoided if possible.

The analysis of variance is computed without difficulty. The method of Appendix III can be used, but it is simpler to proceed as for the completely randomized case, as given in Section 11, but to calculate a sum of squared deviations for blocks in exactly the same way as that used for treatments, i.e. if the block totals are B_{I}, B_{II}, etc, each based on k plots, the sum of squared deviations for blocks is

$$(B_{\mathrm{I}}^2+B_{\mathrm{II}}^2+\ldots \text{etc.})/k-G^2/n$$

where n is the total number of plots.

The analysis of variance is then constructed thus:

Source of variation	*Degrees of freedom*
Blocks	$b-1$
Treatments	$t-1$
Error (by difference)	$(b-1)(t-1)$
Total	$bt-1$

where b is the number of blocks and t is the number of treatments. Because the design is orthogonal, the independent estimation of the sum of squared deviations for blocks and treatments is in order.

22. Latin Squares

In a Latin square the plots are disposed in rows and columns and the treatments are applied so that each occurs once and only once in each row and in each column. For this to be possible, the rows, columns and treatments must all be equally numerous. Figure IV will serve as an illustration. Latin squares have been studied since the eighteenth century (Euler, 1782) and have long been used as an experimental design (Fisher, 1934).

It will readily be seen that the design is based on an orthogonal three-way classification of rows, columns and treatments. Thus all the columns are constituted alike in respect of

D	B	E	C	A
A	D	C	B	E
B	E	D	A	C
E	C	A	D	B
C	A	B	E	D

FIGURE IV. Example of a 5×5 Latin square.

the representation of rows and treatments. Similarly, the rows and treatments are each constituted alike in respect of the remaining classifications.

Latin squares have one notable advantage over randomized blocks in often permitting the trees on the outside of the experimental area to be used equally with those inside. Suppose that the four faces of the trial are different, one perhaps facing a wood, another being exposed to the prevailing wind without shelter and so on. These differences are associated with rows and columns, the effects of which are to be eliminated anyway. Consequently there is no need to waste land with external guard rows as would be necessary with most other designs. It is assumed (i) that each headland is homogeneous, i.e. the wood, or whatever it may be, extends the whole length of the headland, and (ii) that the effect on a corner tree is the sum of the effects of the two headlands meeting there. It is assumed also (iii) that the treatments will behave relatively in much the same way whether they are compared using trees on the inside or the outside of the area.

Latin squares are fairly robust. Thus the loss of data can be dealt with quite easily (see Section 83), as can the loss of a complete row, column or treatment. The latter case has been studied in some detail (Yates, 1936a) the method of analysing results being also derivable from Appendix III.

Latin squares are more flexible than is sometimes supposed, but that will be considered in the next Section. They do, however, suffer from the disadvantage of being useful for only a limited number of treatments, say four to seven. Smaller squares do not give enough error

(a)

D	A	C	B	C	D	B	A
B	C	A	D	D	C	A	B
C	B	D	A	B	A	C	D
A	D	B	C	A	B	D	C

(b)

B	B	D	A	C	C	A	D
A	D	B	B	D	A	C	C
C	A	C	D	B	D	B	A
D	C	A	C	A	B	D	B

FIGURE V. Examples of double 4×4 Latin squares. (a) No tying. (b) Tied.

degrees of freedom, though a 3×3 square used in duplicate or triplicate often provides a good design. In larger squares the control of environmental effects is attempted by rows and columns that are so long and narrow as to be virtually useless. The plots themselves need not be square. The chief restriction upon the use of Latin squares is the requirement that rows, columns and treatments must all be equal in number.

The analyses of variance for a $k\times k$ Latin square is constructed thus:

Source of variation	*Degrees of freedom*
Rows	$k-1$
Columns	$k-1$
Treatments	$k-1$
Error	$(k-1)(k-2)$
Total	k^2-1

By a multiple Latin square is meant two or more Latin squares, not necessarily adjacent, having the same treatments and analysed together. Sometimes it is possible for them to stand side by side and for the rows to run through, a state of affairs well described by saying they are "tied". Thus in Figure V(b) the squares are tied but not in Figure V(a), the difference being that the columns are all randomized together in the one case but in two sets of four in the other.

In the main, these designs have the same advantages and disadvantages as simple Latin squares, the important difference being the greater number of replicates. As to tying, if the squares do not lie side by side it is impossible, but generally it is advantageous on account of the additional degrees of freedom for error. With untied squares analysis can be a little tricky, because the rows and columns are randomized within the squares, which are only part of the whole experimental area. The correct analysis for b untied $k\times k$ squares is set out below. (The "interaction", i.e. that between squares and treatments, will be defined later in Section 24. For the moment it is enough to note that it represents the extent to which the treatments have different effects in different squares.)

Source of variation	*Degrees of freedom*
Squares	$b-1$
Rows within squares	$b(k-1)$
Columns within squares	$b(k-1)$
Treatments	$k-1$
Interaction, squares × treatments	$(b-1)(k-1)$
Error, by difference	$b(k-1)(k-2)$
Total	bk^2-1

If the squares are close together, the treatments can be expected to act relative to one another in the same way in all squares so the interaction can be merged with error. Incidentally, the error sum of squares can be checked by working out the error for each Latin square separately. Their sum should equal the error obtained for the untied multiple square.

23. The Changing of Treatments in a Latin Square

With randomized blocks the possible future uses are very wide. With Latin squares they are more narrow and it will be convenient to discuss them here.

It should first be emphasized that a change of treatments in the way to be described is permissible only where there is no likelihood of the two sets of treatments interacting. Thus,

if the original treatments have left some trees growing well and others nearly stationary in growth, it is not to be expected that they will all react alike to a subsequent set of pruning treatments and it would be most unwise to effect such a change of purpose. Of course, if a period were to elapse during which the trees were treated alike or if efforts had been made to even them up by reversing treatments, the position might become very different. On the other hand, the fact that the original treatments might be expected to leave a permanent residual effect need not affect the differences between the following treatments. It would be of no importance because the differences will be eliminated in the analysis of variance.

The subject has been studied exhaustively for squares of size 4×4, 5×5 and 6×6 with conclusions important for all who design trials with long-lived plants. Before these results can be presented, however, it is necessary to explain what is meant by a "transformation set".

To take an example, all 4×4 squares can be generated from one or other of these two squares

I					II				
	A	B	C	D		A	B	C	D
	B	A	D	C		B	A	D	C
	C	D	B	A		C	D	A	B
	D	C	A	B		D	C	B	A

by permutation of rows, columns and letters. Now it happens that squares generated from the second, i.e. those of Transformation Set II, are more flexible than the others (Finney, 1945a). Further, restriction of randomization to secure a square from a chosen transformation set does not bias the tests of significance (Fisher, 1935). Consequently, an experimenter with perennial plants who foresees the possibility, however remote it may be, of losing interest in his original set of treatments and wanting to replace them by a new set is advised to abandon the usual method of selecting Latin squares (Fisher and Yates, 1938). Instead he should take Square II, permute its rows at random, then permute at random the columns of the resulting square and then assign the letters at random to his four treatments.

Supposing he does so, what alternatives has he if he does want to change? Writing small letters for the new treatments and capitals for the original ones, the most interesting possibility is a Graeco-Latin square, thus:

Aa	Bb	Cc	Dd
Bd	Ac	Db	Ca
Cb	Da	Ad	Bc
Dc	Cd	Ba	Ab

Each new treatment occurs once in each row, in each column and upon each original treatment. There are four such treatments, so the "partition" will be said to be of the type (1^4). If the scheme had been

Aa	Ba	Cc	Dd
Bc	Ad	Da	Ca
Cd	Dc	Aa	Ba
Da	Ca	Bd	Ac

two new treatments (c and d) occur once on each of the original classes and one (a) occurs twice. The partition is thus of the type $(1^2, 2^1)$, usually written $(1^2, 2)$.

The 4×4 squares of Transformation Set II permit partitions of all possible types, namely, (1^4), $(1^2, 2)(2^2)$ and $(1, 3)$.

Similarly, if 5×5 squares are chosen from those of Transformation Set II, i.e. by permutation of rows, columns and letters starting from

A	B	C	D	E
B	C	E	A	D
C	E	D	B	A
D	A	B	E	C
E	D	A	C	B

they will permit of all possible types of partition namely, (1^5), $(1^3, 2)$, $(1^2, 3)(1, 2^2)$, $(2, 3)$ and $(1, 4)$ (Finney, 1946a).

Unfortunately there is no 6×6 Latin square that will permit of a partition of type (1^6), the most useful kind (Fisher and Yates, 1934). However, those of Transformation Set X† permit all the other possibilities (Finney, 1946b, c), $(1^4, 2)$, $(1^3, 3)$, $(1^2, 2^2)$, $(1^2, 4)$, $(1, 2, 3)$, $(1, 5)$, (2^3), $(2, 4)$ and (3^2). Such squares are those generated by permutation of rows, columns and letters, starting from

A	B	C	D	E	F
B	C	F	A	D	E
C	F	B	E	A	D
D	A	E	B	F	C
E	D	A	F	C	B
F	E	D	C	B	A

Apart from the work of Norton (1939) on the 7×7 squares, there is little known about larger squares than those considered here. A 7×7 square giving a transformation set of maximum flexibility is the following:

A	B	C	D	E	F	G
B	F	E	G	C	A	D
C	D	A	E	B	G	F
D	C	G	A	F	E	B
E	G	B	F	A	D	C
F	A	D	C	G	B	E
G	E	F	B	D	C	A

Fisher and Yates (1938) cite "orthogonal squares" of size 8×8 and 9×9, which may be used as a starting point for randomization to give transformation sets capable of all possible partitions, including Graeco-Latin squares. Bose, Shrikhande and Parker (1960) provide possible 10×10 squares.

All the designs considered here are derived from orthogonal four-way classifications (rows, columns, original treatments and new treatments). Consequently they cause little difficulty in the analysis of data. Simplest of all are the Graeco-Latin squares, i.e. designs of the type (1^n), in which all the new treatments are equally replicated.

The randomization of these designs, if it is to be done correctly, requires some thought. If a Graeco-Latin square, to take an example, is written down and then its rows and columns are permuted at random, the two sets of treatments being allocated at random to their respective sets of letters, the resulting analysis of variance gives unbiased tests in respect of both original and new treatments. The problem here is to convert an existing Latin square into a Graeco-Latin square or similar design. Merely to fill in the letters for the new treatments as well as may be is not enough, because it might lead to unconscious groupings or dispersal of them over the experimental area. Instead, it would seem better to return to the basic square from which the existing trial had been derived and to add the new treatments to that. Then it only remains to re-enact the permutations previously gone through to produce again the original design but this time with a further classification added, finally allocating the new

† Strictly there are 22 transformation sets of 6×6 squares, but, as a matter of convenience, conjugates are here ignored. Numbering thus follows that given by Fisher and Yates (1934, 1938).

treatments at random to the letters representing them. In order to be able to do this conveniently, it is a good plan to keep a note how each Latin square was derived from the basic square.

The emphasis here has been upon orthogonal partitions, but non-orthogonal ones are possible as well. Consideration of them will be left to Chapter 3, when they will be dealt with alongside the non-orthogonal designs that can be based on randomized blocks.

Given an orthogonal partition in which H new treatments are added to a $K \times K$ Latin square, there is no difficulty about constructing the analysis of variance, which should go like this:

Source of variation	*Degrees of freedom*
Rows	$K-1$
Columns	$K-1$
Original treatments	$K-1$
New treatments	$H-1$
Error	$K^2-3K-h+3$
Total	K^2-1

The line for new treatments may present a problem if the treatments are unequally replicated. All that is needed is to square each treatment total of data and to divide by its replication. The quotients are added and finally G^2/K^2 is subtracted, G being the grand total of all data.

24. Factorial Designs

In this Chapter so far there has been an implication that the treatments form a single series, say, of four pruning methods or of five varieties, but that is not necessarily the case. For example, an experimenter might want to try out his pruning methods on a range of varieties or he might want to study several fertilizer treatments both on clean cultivated soil and on grass.

A note on nomenclature may here be useful. If, for example, three levels of nitrogen are used in conjunction with three levels of potash on two varieties of rootstock, there are three "factors" (nitrogen, potash and rootstocks), two of them at three "levels" and one at two. That can be expressed by saying that the design has $2.3^2 (= 18)$ "treatment combinations".

All this is something apart from the design as such, which concerns the way in which local control of environment is effected by the formation of blocks and by the way in which treatments, or treatment-combinations, are allocated to them. A particular design having been adopted, there is one corresponding way of partitioning the total sum of squared deviations and total degrees of freedom among the various blocking systems, the treatments and the error. What is involved here is a further partitioning, this time of the treatment sum of squared deviations and degrees of freedom to take account of the treatment structure. Hence, any of the designs already considered could have been used with a factorial set of treatments. However, the subject is introduced here because some of the designs that follow assume a factorial structure of treatments.

If Factor A has a levels and B has b, there will be $(ab-1)$ degrees of freedom between the treatment combinations. If now factor B is ignored to give a "main effect" of A, there will be $(a-1)$. Similarly, there will be $(b-1)$ if factor A is ignored. What has happened to the

$$(ab-1)-(a-1)-(b-1) = (a-1)(b-1)$$

degrees of freedom that arise from the treatment-combinations but appear not to be associated with either factor? The answer is they relate to the "interaction" of A and B, written $A \times B$, i.e. to effects that do not arise from either factor alone but from the two working in concert.

Thus, in the first edition data were presented from a pear pruning trial, in which the following treatment means were obtained. They represent crop per tree over a ten-year period.

Few leaders,	Hard pruning	492 lb
Few leaders,	Light pruning	565 lb
Many leaders,	Hard pruning	526 lb
Many leaders,	Light pruning	641 lb

Each mean is based upon forty trees. The error variance was 9096 with 133 degrees of freedom.

It will be seen that the "particular" effect of the number of leaders is 34 lb when pruning is hard and 76 lb when it is light. Put the other way, the particular effect of the degree of pruning is 73 lb if there are few leaders and 115 lb if there are many. It might well appear that the two factors were not acting independently, but the mean square for the interaction is only 4656 so F is actually less than one, which illustrates the little-recognized fact that interactions are not shown easily. However, there is no prior reason for thinking that the factors cannot be independent and no evidence that they do depend upon one another for their effect, so the next stage is to look at the "main effects". That for number of leaders is $\frac{1}{2}(34+76) = 55$ with F equal to 18·39***, while that for degree of pruning is $\frac{1}{2}(73+115) = 94$ with F equal to 47·48***, so here results are clear-cut.

Suppose, however, that there had been a marked interaction that was clearly significant. In that case it would have been wrong to have proceeded to a study of main effects because they would have made no sense. If the effect of harder pruning depends upon the number of leaders, it is no good trying to evaluate how large it is without specifying how many leaders are being kept. All that could or should be done is to give either the whole table or to confine discussion to the particular effects.

An exception occurs with the so-called "mixed model" where the levels of one of the factors are effectively a random selection of the possibilities. Thus, if a spray experiment is carried out with several replicates each year on fresh material, it could well happen that an interaction between years and treatments should be assumed, but that would not imply that nothing could be said about treatment differences, which might be much the same in each season, though not exactly so. Carrying out an F-test on the mean squares for the main effect of treatments and the interaction of treatments and years, it might appear that what was constant in the treatment effects outweighed the ephemeral. If the years could be considered to be representative, not all wet or all dry or all cold or all warm, it would be reasonable to assert that the treatments had had certain effects which persisted irrespective of season, though not necessarily to exactly the same degree. Essentially the same argument applies with a randomized block design. The error is in fact the interaction of blocks and treatments. If the treatments show much the same effect irrespective of blocks, differences between them are accounted "significant"; on the other hand, if they behave quite differently in different blocks, no conclusion about them is established.

The mixed model is not easy to use if there is an independent estimate of error. Then, if there is no evidence of an interaction the independent estimate is used, but if there proves to be one then that interaction becomes the error. If the interaction is clearly not significant, someone will be found who wants it and error to be merged to obtain a better estimate of error with more degrees of freedom, but that amounts to merging the interaction when it is small and not doing so when it is large, which necessarily leads to bias. A more difficult situation arises when the interaction appears to be significant relative to the error and some believe in it and others do not. If the experimenter believes in it when he reports it, his reader may disagree.

It is possible for several factors to interact. Thus, if there are three, A, B and C, with a, b and c levels respectively, out of $(abc-1)$ degrees of freedom, $(a-1)+(b-1)+(c-1)$ can be attributed to main effects and $(b-1)(c-1)+(a-1)(c-1)+(a-1)(b-1)$ to two-factor

interactions, leaving $(a-1)(b-1)(c-1)$ for the three-factor interaction. If it should appear significant, there should not ordinarily be any further study of main effects and interactions that depend upon it, the three-way table being the best way of setting out what was shown.

However, even when a factorial partition is possible it is not always appropriate. For example, an experimenter might have a range of varieties, each virus-infected to some degree, and he might have two clones of each, one that had been heat-treated to make it virus-free and the other not so treated. After an experiment on those $2C$ clones, he might unthinkingly partition the $(2C-1)$ degrees of freedom for treatments thus:

Varieties	$C-1$
Virus	1
Interaction	$C-1$

That would be wrong. He can assume that the virus contents of his various varieties will be different and anyway varieties have their individual reactions to virus infection, so the interaction can be taken for granted. Also, where there is an interaction, it makes no sense to look at the main effects, since a warning has been given that the effect of each factor depends upon the level of the other. The correct partition is

Varieties	$C-1$
Virus in Variety 1	1
Virus in Variety 2	1
.....	
Virus in Variety C	1

There is still no meaning to be ascribed to "Varieties" as such, but the rest of the partition makes good sense.

In order to make partition of the treatment effects possible there is a certain pressure on experimenters to build factorial structure into their experiments and in general they should do so. Many an investigation has been weakened by the omission of a treatment-combination that would complete the factorial set. For example, a study is to be made of two ways of grafting, X and Y, at two different heights and it is proposed to graft 12 trees each of (A) X high, (B) X low and (C) Y high. Asked what has happened to (D) Y low, the experimenter replies that the effect of high grafting versus low should be unaffected by method, so D is unnecessary. Nevertheless, it would be better to have nine trees on each of the four treatments. There would then be 18 trees on each side of the comparison of methods (A and B compared with C and D) instead of 12 in the original scheme (A compared with C, with B irrelevant) and similarly for the comparison of heights. Also, if the choice of method should depend upon the height at which it is proposed to graft, i.e. if there *is* an interaction, there would be a means of detecting it.

Nevertheless, an insistence upon always having a complete factorial set can become a fetish. Certain treatment-combinations may be little better than absurd. For example, if a weak-growing rootstock will almost certainly die if planted with a vigorous grass mixture or if a variety will bear no fruit if pruned in a certain way, the trees might just as well be omitted in the first place as left to languish uselessly. Also, as has been explained, there are occasions when a factorial partition is not appropriate even when it is possible. Some useful papers are those of Wishart (1938), Finney (1948) and Williams (1952).

25. Split-plots

In essence, a split-plot design is one in which the plots of one experiment are used as the blocks of another (Yates, 1937). Thus, a cover crop trial might be laid down with four-tree

plots, which might be used as blocks for a trial of four pruning treatments, one tree of each plot being assigned at random to each of the pruning treatments. This process of splitting can be continued almost indefinitely. Thus, the plots of the first trial having been split into sub-plots for a second set of treatments, those sub-plots, if large enough, can be further split into sub^2-plots for yet a third set of treatments and these sub^2-plots can be again split into sub^3-plots and so on.

Split-plots are used for three main reasons. For one thing, they enable emphasis to be put where it is needed. Thus, an experimenter might want to test out a range of six orange scion varieties and might want to use a range of four contrasting rootstocks, not because he wants to study the rootstocks themselves, their general characteristics being well known, but because he wants to detect any interaction there may be of stock and scion. If he lays down a series of randomized blocks, each of 24 trees, one for each combination, his control of environmental variation is likely to be poor on account of the large block size. Instead, he might prefer to divide blocks of 24 trees into four plots of six, one plot for each rootstock, and to divide each such plot into six one-tree sub-plots for the scion varieties. The varieties are now compared in one-tree units within areas of six trees, instead of areas of 24 trees as formerly, so their comparison is improved. The rootstocks, on the other hand, are applied to a certain number of trees in groups of six instead of well dispersed, so their comparison is worse. In fact, emphasis has been placed where it is wanted. However, as will appear below, split-plot designs have certain technical disadvantages.

Another reason for using split-plots is convenience in the field. Some treatments (e.g. differential soil management, sprays and systems of planting) cannot reasonably be applied to small areas, whereas others (e.g. varieties, rootstocks and pruning methods) can be applied to individual trees. Often the most practical way of designing an experiment is to let each factor be applied to a unit that suits it. Thus, if an investigation is to be carried out into methods of soil management on a range of varieties, it might well be best to make up plots that contain each of the varieties once and to apply the differential soil treatments to them. Indeed, any other approach might be absurd; there is a lower limit to the areas that can be assigned different cultivation treatments, while varieties may need to be dispersed among each other to secure good pollination.

Sometimes the two considerations can conflict. Thus, in the experiment just described, if the main purpose were to study the varieties under the range of soil treatments that they might encounter in commerce, the applying of cultural treatments to main plots and of varieties to sub-plots would cause no problems. The better comparison would be of varieties, as desired. Also, if there were an interaction of varieties and cultural methods, there would be no difficulty about working out the standard error of differences between varieties. If, however, the main

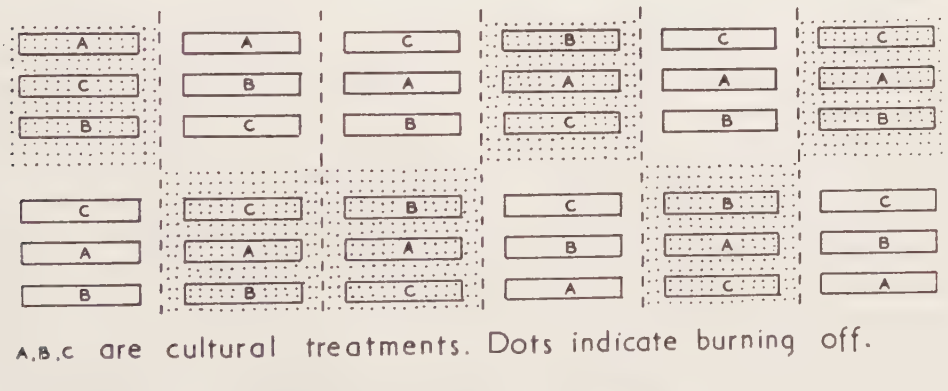

FIGURE VI. Example of a split-plot design.

purpose were to study methods of soil management and to do so on a range of varieties, several difficulties arise. One is the better comparison of the less important factor. Another is the calculation of a standard error for differences of soil treatments, the variety being fixed. A method is given in Section 72 but it involves complications that are better avoided.

A third occasion for splitting plots is the addition of a new set of treatments, the original set remaining of interest. Thus, the East Malling strawberry trial illustrated in Figure VI was originally intended to compare three cultural methods, A, B and C, but the question arose later whether or not to burn off the plants. That was investigated by grouping the blocks of cultural treatments into pairs and burning off one but not the other, thus converting the trial into one in split-plots, as shown. (Actually, there was a restriction of randomization here also—quite a valid one this time provided it is taken into account when analysing the data—since in each row of the original blocks, three were burnt and three not.) Another way to use split-plots for the addition of treatments is to divide the original plots instead of grouping the original blocks.

With a split-plot design the analysis of variance is readily constructed. Thus, if the main plot experiment is in b randomized blocks with m treatments and there are s sub-treatments, it goes like this:

Source of variation	*Degrees of freedom*
Blocks	$b-1$
Main treatments (M)	$m-1$
Error (i)	$(b-1)(m-1)$
Total (i)	$bm-1$
Sub-treatments (S)	$s-1$
Interaction (M × S)	$(m-1)(s-1)$
Error (ii)	$m(b-1)(s-1)$
Total (ii)	$bms-1$

If the main plot experiment had been of some other design, as in Figure VI, the first (or upper) analysis would have been of the kind appropriate to the design, but the second would have been unaltered, b being then defined as the number of plots of each main treatment.

26. Criss-cross Designs

Sometimes it is a help if the treatments can be applied in strips that terminate on a headland. For example, cultivation treatments cause trouble if there is no place for a tractor and its implement to turn and a similar situation arises when varieties need to be planted in complete rows for reasons of pollination or equalization of competition. In these circumstances the criss-cross design can be very useful.

In this design, illustrated in Figure VII, the experimental area is split up into a number of rectangles, b in number, which are not necessarily orientated alike and not necessarily contiguous, so alleys may be left between them if need be. Suppose that there are t treatments in one set and u in the other, then each rectangle must be divided into tu plots, t by u, so that one set of treatments can be randomized on to the columns and the other on to the rows of the rectangles.

The analysis of data from such a design is quite easy. There are in all btu plots, arising from all combinations of b rectangles of t "first treatments" and u "second treatments". Consequently, the $(btu-1)$ degrees of freedom can be split up as in a factorial group of treatments and the sum of square deviations likewise, with the following result:

Source of variation	*Degrees of freedom*
a. Rectangles	$b-1$
b. Treatments I	$t-1$
c. Treatments II	$u-1$
d. Treatments I × Treatments II	$(t-1)(u-1)$
e. Treatments I × Rectangles	$(b-1)(t-1)$
f. Treatments II × Rectangles	$(b-1)(u-1)$
g. Treatments I × Treatments II × Rectangles	$(b-1)(t-1)(u-1)$
Total	$btu-1$

Each of the sources of variation, b, c and d, may be tested for significance by using its interaction with rectangles as error, i.e. by using e, f and g respectively. There are thus three errors in an experiment of this kind.

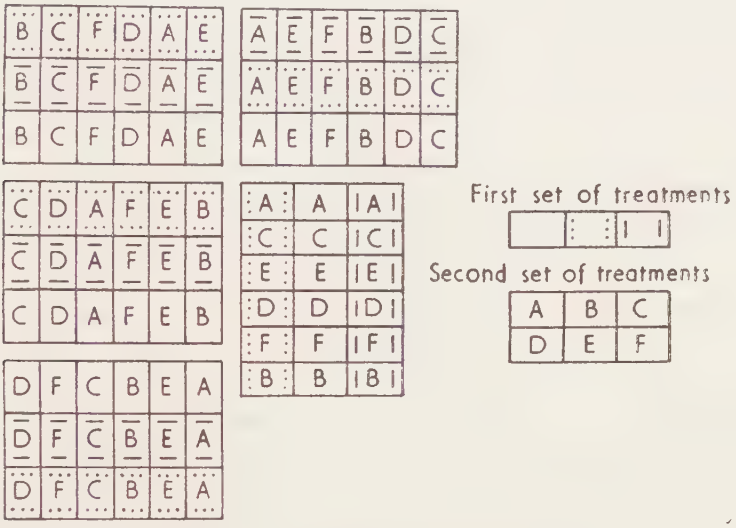

FIGURE VII. Example of a criss-cross design.

27. Unrandomized factors

The need for randomization is nearly absolute if an experiment is to be soundly based, but where treatments have factorial structure it is permissible to randomize only those factors actually under study.

To appreciate the point of doing so, it is helpful to recall the nature of experimental error as given by an analysis of variance. Basically, of course, it represents the residue of variation that cannot be accounted for by blocks or treatments, but often it has the form of an interaction with treatments. That is explicitly so for randomized blocks and criss-cross designs, though true only with qualifications for some others, such as Latin squares, but as a generalization it will serve. In fact, the experiment evaluates a contrast of treatments several times, say, once in each block. Only if what is constant in the difference (as given by the variation between treatment means) exceeds what is arbitrary (as given by the interaction of treatments with blocks) does the experimenter consider that he has established an effect of his treatments that can be relied upon in general.

Thus, in a split-plot design it is quite permissible to think of it in this form:

Source of variation	*Degrees of freedom*
a. Blocks (B)	$b-1$
b. Main treatments (M)	$m-1$
c. Error (i) (B × M)	$(b-1)(m-1)$
Total (i)	$bm-1$
d. Sub-treatments (S)	$s-1$
e. Interaction (M × S)	$(m-1)(s-1)$
f. Error (ii) = B × S	$(b-1)(s-1)$
g. + B × M × S	$(b-1)(m-1)(s-1)$
Total (ii)	$bms-1$

If now, the main plot treatments have been introduced only for the interaction, there is no necessity to randomize them. By disposing them systematically, e.g. in strips, line *b* is vitiated and so is line *c* to compensate; consequently, no use should be made of the first analysis, but in the circumstances it is not going to be needed. Similarly, if the sub-treatments were introduced for the same purpose e.g. they represent a range of diverse varieties, they could be disposed in any convenient manner, e.g. to promote pollination, thus vitiating lines *d* and *f*, but leaving a valid *F*-test between lines *e* and *g*.

Again, in the case of a criss-cross design, one of the factors can be disposed systematically. Its mean square can no longer be compared with that for its interaction with blocks, since both will be vitiated, but the rest of the analysis is all right. Similar situations arise with other designs.

CHAPTER 3

SOME MORE COMPLICATED DESIGNS

"A mighty maze! but not without a plan".
ALEXANDER POPE, *Essay on Man*, i, 6.

30. Non-orthogonal Designs in General

The designs considered above have all been orthogonal, though not always in the sense of the rather restricted description given in Section 21. Before proceeding further, a fuller definition may be attempted: Two classifications are mutually orthogonal if the various groups of plots formed by one classification are composed of the same proportionate number of plots of the other. Where all pairs of classifications are mutually orthogonal, the whole design is said to be orthogonal.

Thus, in the trial shown in Figure VIII there are two classifications, namely, blocks and treatments, so, if they are orthogonal, the whole design is. First, it will be seen that each block is made up in the same way, one-sixth of its plots receiving Treatment A, one-third Treatment B and one-half Treatment C. Further, the plots of each treatment are all divided in the same way, one-quarter occurring in each block.

A	C	C	C	B	B
B	C	B	A	C	C
C	C	C	B	C	C
B	A	B	C	A	B

FIGURE VIII. Example of an orthogonal design.

The basic advantage of orthogonality is simply the independent way in which the effects of the various classifications may be sorted out. Thus, in Figure VIII, if one of the blocks happens to be situated on bad soil, the same proportion of plots will be affected in each treatment, while if one of the treatments proves more effective than the others, all the blocks will be affected to the same extent. With a non-orthogonal design, such as that in Figure IX, that is not so, e.g. treatment A does not occur in Block III. If the block is on worse or better soil than average, some of the treatment means will need adjustment but not others. Likewise, if a treatment differs from the rest, some but not all of the block means will require adjustment.

Where adjustment is needed there is necessarily an increase in the standard error of the mean, because the adjustment is itself an estimated quantity derived from the data and therefore itself subject to error. The expressions given in Section 11 referred to designs without blocks. They can also be used for orthogonal designs where the blocks, though present, do not affect treatment means. With non-orthogonal designs, however, there must be a loss of "efficiency". An example will show how it is measured. Suppose that there are two means,

A	F	G	B	E	A	D
G	A	F	E	F	B	B
C	D	C	D	C	E	A
E	G	B	G	D	F	C
I	II	III	IV	V	VI	VII

FIGURE IX. Example of a balanced incomplete block design.

each based on eight plots, then the standard error of their difference should, according to Section 11, be the square root of (0·25 × error variance). If as a result of adjustment it is in fact the square root of (0·40 × error variance) the "efficiency factor" is taken to be $\frac{5}{8}$, i.e. 0·25 divided by 0·40. It is as if each plot had only $\frac{5}{8}$ of its value for that particular purpose.

A further difficulty with non-orthogonal designs arises from complications when the treatment line in the analysis of variance requires to be partitioned on account of factorial structure or other patterns among the treatments. A group of degrees of freedom can be partitioned freely only when all contrasts have the same efficiency factor (Caliński, 1971, Pearce, Caliński and Marshall, 1974). With orthogonal designs, where all contrasts have an efficiency of one, there is no problem; with many non-orthogonal designs the problems are not serious, but the point needs to be watched.

Non-orthogonality is not to be strenuously avoided. In fact, it often arises from mishaps that can be easily dealt with (Chapter 8) and the difficulties can be even less if it is introduced deliberately, but there are certain other disadvantages. For one thing, it calls for more sophisticated computer programs, which may not be available. The complications are at their worst when covariance adjustments have to be made or missing values fitted and both contingencies occur with some frequency with perennial plants. For another, the need to adjust treatment means bothers some people quite a lot; they work out averages from the data sheet and then the computer produces something different. The objection seems rather unreasonable when other adjustments, e.g. those arising from covariance (Chapter 4) or from transformations (Section 73), are accepted with equanimity, but it continues to be made.

Sometimes only certain treatments are orthogonal to blocks. Where that occurs the orthogonal treatments give means that require no adjustment nor do differences between them, while standard errors that relate to them can be derived as for an orthogonal design. Thus, if an experiment is designed in randomized blocks but an error in planting arises so that Treatment A occurs twice in Block I but not at all in Block II, while Treatment B occurs twice in Block II but not in Block I, the other treatments will still be orthogonal to blocks.

Further, even though two treatments are non-orthogonal it is still possible for the difference of their means to require no adjustment because the two means individually are adjusted to the same extent. That happens when a constant ratio is maintained in the blocks between the number of plots allocated to the two treatments, which are then said to be "proportionate" (Pearce, 1971). To take the design in Figure X as an example, in each block there are twice as many plots of Treatment B as there are of A and that would remain true whatever changes

B	A	A	C	C	D	E	C
E	D	B	A	D	E	A	D
C	A	E	D	C	E	A	B

FIGURE X. Example of design with two proportionate treatments (A and B).

of plan led to an allocation of the remaining plots among treatments C, D, E, etc. Consequently, the difference of means of treatments A and B would still require no adjustment and its standard error could still be derived from the error variance as in Section 11.

To avoid confusion it should be emphasized that two ratios are involved. For orthogonality of a treatment to blocks, it is required that the number of plots allocated to the treatment in each block should bear a constant ratio to *the block size*. For two treatments to be proportionate, it is required that the number of plots allocated to one in each block should bear a constant ratio to *the number allocated to the other*. The two are not necessarily the same, especially if blocks are of unequal size.

The possibility of using a non-orthogonal design can be extremely welcome. Sometimes the allocation of plots to blocks is virtually fixed by events. There are, say, eight plants from one source and twelve from another. Blocks of five plots each may be desirable but there is no reasonable way of forming them.

One occasion for restriction of block size is the use of the guard rows of an existing trial for a subsidiary experiment. Thus, in Figure XI it is supposed that plots of an extant experiment consist of four trees arranged in a row, separated by single guards. The guard trees indicated by X fall into groups of four, all the trees of a group being treated alike as far as the main trial is concerned, though each of them has a different treatment on its two sides. If the treatments of the proposed subsidiary trial are unlikely to interact with those already being applied, those groups of four trees make a natural system of blocks for the new experiment, but what if there are to be more than four treatments? (Of course, no treatments should be applied to the guard rows that might disturb the original trial.)

In such cases it is a relief to have available non-orthogonal designs because orthogonal ones clearly will not do. As to difficulties of partition, they can be given too much importance. Strictly speaking the partitions required in the treatment component of the analysis of variance are determined by the nature of the treatments and their relationship one to another. In fact, it is rarely necessary to be able to partition *freely*. Admittedly, however, it is sometimes

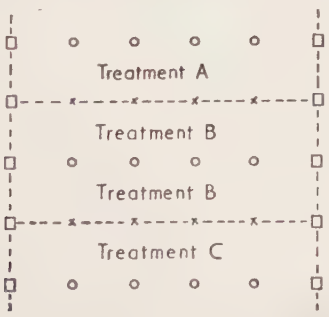

FIGURE XI. Use of guard trees for a subsidiary trial.

a problem to devise a non-orthogonal design to fit a given blocking system that will give maximum efficiency to the contrasts of greatest interest among the treatments and allow of all needful partitions. It is worth noting that if a group of treatments are all orthogonal to blocks or all are mutually proportionate, freedom of partition among them is retained (Pearce, 1971).

Statistically it matters little whether the blocks are "incomplete", i.e. too small to take a complete set of treatments, or "extended", i.e. larger than is required for completeness. In general, incomplete blocks are the more uniform within themselves but extended blocks mostly give higher efficiency factors, so both have their uses. What is important is to accept any natural grouping of plots and to form blocks on the basis of it, avoiding artificiality (Pearce, 1964). Sometimes there is no natural blocking system and it is possible to divide the area at will. When that is so, a complete set of treatments to each block is the norm and other block sizes should be adopted only with circumspection.

Latterly there has been some interest in extended blocks because of the possibility of obtaining both a pure error and an interaction of blocks and treatments (John, 1962, 1963; Cornell, 1974). Thus, in the following design:

I	A	A	B	C	D
II	A	B	B	C	D
III	A	B	C	C	D
IV	A	B	C	D	D

the analysis of variance would ordinarily be constructed thus:

Source of variation	*Degrees of freedom*
Blocks	3
Treatments	3
Error	13
Total	19

Now, it is true that the error, so-called, can be partitioned into an interaction of blocks and treatments with nine degrees of freedom and a pure error with four, derived from the comparison of a pair of similar plots within each of the four blocks (Trail and Weeks, 1973). However, to argue thus is to invite all the complications and frustrations of the mixed model, as discussed in Section 36, though they must be accepted if the situation requires. If the experimenter really believes that there could be an interaction of blocks and treatments he should adopt it as his true error. After all, that is what constitutes experimental error, i.e. the consistency of a contrast or other quantity when repeated in a number of representative circumstances. If there is genuinely an interaction in the sense that treatments behave relatively differently in different blocks, the pure error is a distracting irrelevance. If, on the other hand, it is supposed that the underlying relative behaviour of the treatments is much the same anywhere in the experimental area but individual plots are affected differently by local variation or random errors, then the pure error is as valid as the interaction and the two should be pooled. There are no circumstances in which the pure error should be adopted alone, though some appear to advocate it (e.g. Cornell, 1974). It is, of course, useful as a means of judging the significance of the interaction supposing the experimenter to be in an enquiring mood about the source and nature of the error he has to deal with. Clarke (1963) has drawn attention to the advantages of non-orthogonal extended designs in conserving error degrees of freedom.

31. Balanced Incomplete Blocks and Other Totally Balanced Designs

The useful design known as "Balanced Incomplete Blocks" (Yates, 1936b) is of interest in showing clearly the principle of "balance" which is desirable whenever orthogonality is

unattainable. Thus, the trial in Figure IX is laid out in seven blocks ($b = 7$) but they are "incomplete", only four ($k = 4$) treatments out of seven ($v = 7$) occurring in each. Examination will show that each treatment "occurs" an equal number of times in the trial. Thus treatment A occurs four times ($r = 4$), in Blocks I, II, VI and VII, as does treatment B in Blocks III, IV, VI and VII. Also, each pair of treatments "concurs" an equal number of times in the trial. For example, treatments A and B concur together in Blocks VI and VII, and treatments A and C likewise concur twice, i.e. in Blocks I and VII ($\lambda = 2$). It is this equalizing both of occurrences and concurrences that is the essence of balancing. Possible designs have been listed by Fisher and Yates (1938). It may be noted that λ always equals $r(k-1)(v-1)$ and bk always equals rv. It is assumed that all blocks are of equal size.

As has been mentioned, non-orthogonal designs always lose something in precision on account of the adjustments needed to the treatment means. For balanced incomplete blocks, the efficiency is $v\lambda/rk$, or 0·875 in the present instance. This means that the four replicates of the trial in Figure IX would give as much information as $(4 \times 0{\cdot}875)$ or $3\frac{1}{2}$ replicates in randomized blocks.

The usual reason given for accepting this loss of information is that it will be recouped by the smaller size of block diminishing uncontrolled environmental variation, but the argument will not impress the experimenter with trees if he believes that a good part of his

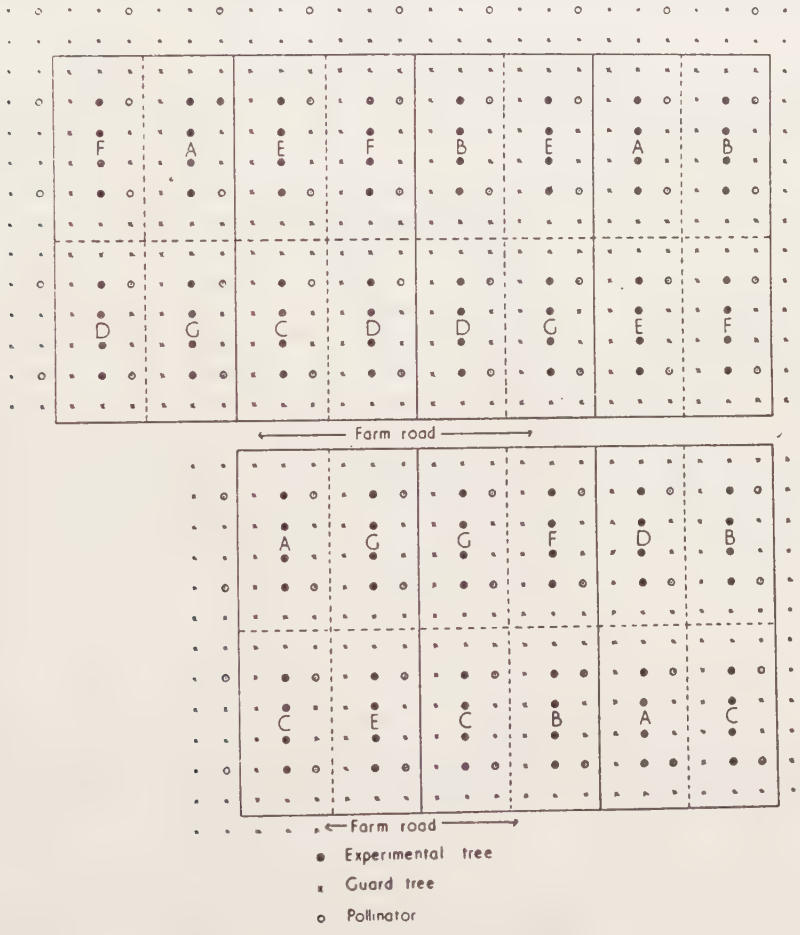

FIGURE XII. Plan of a spraying trial designed in balanced incomplete blocks.

experimental error is due to his plant material rather than to position. However, the design is useful for another reason because blocks of large plants cannot be adapted in shape and size to the experimental area as can blocks of small ones. Thus, in Figure XII, which represents an application of Figure IX to a trial for controlling red spider on apple trees, there is no way of keeping the chosen plot size and fitting in four compact blocks of seven plots for a randomized block design, especially when a large degree of guarding at one end was insisted upon to keep the sprays off some vegetables to be harvested for human consumption soon after the application of treatments. On such occasions, when the block size is restricted, balanced incomplete blocks provide a useful resource.

In balanced designs the treatments are neither orthogonal to blocks nor proportionate to one another, but they do exhibit a third useful property, that of "equivalence". That is to say, if they are permuted the design remains in essentials what it was before. In Balanced Incomplete Blocks all treatments are equally replicated and all pairs of treatments concur equally often, so all are equivalent. Among equivalent treatments partitioning may take place freely. It is possible for all treatments to be equivalent even though the design is not in Balanced Incomplete Blocks, the wider class thus defined being said to possess total balance (Pearce, 1963), or to be of Type T. For example, the block size might be greater than the number of treatments, as in this example:

I	A	B	B	C	C
II	A	A	B	B	C
III	A	A	B	C	C

or the blocks may not even be equal in size, as in this:

I	B	C	
II	A	C	
III	A	B	
IV	A	B	C

In both examples, anything that can be predicated of one treatment can be predicated of any other (e.g. replication) and anything that can be predicated of a pair of treatments can be predicated of any other pair (e.g. the number of concurrences with other treatments in blocks of a given size), so all are equivalent. That being so, the effective replication for the study of contrasts between them is derived thus: Two treatments, say A and B, are chosen arbitrarily, then each block is taken in turn and the number of plots allocated to A is multiplied by the number of plots allocated to B and the product is divided by the number of plots in the block, the results being added. For the first of the examples given above the sum, p, is

$$\tfrac{1}{5}(1\times 2)+\tfrac{1}{5}(2\times 1)+\tfrac{1}{5}(2\times 2)=\tfrac{8}{5};$$

for the second it is

$$\tfrac{1}{2}(0\times 1)+\tfrac{1}{2}(1\times 0)+\tfrac{1}{2}(1\times 1)+\tfrac{1}{3}(1\times 1)=\tfrac{5}{6}$$

The quantities, p, are then multiplied by the number of treatments, three in each instance, to obtain the effective replication. For a first example it proves to be 4·8, the actual replication being 5, so little will have been lost; for the second example it is 2·5, appreciably less than 3, so the design is less satisfactory, though it would be foolish to expect too much of small blocks.

If there is any doubt whether treatments are really equivalent, a useful check is provided by carrying out the above calculation for all pairs. Obviously they are not equivalent if replications are unequal, but concurrences are more difficult to judge. In the two examples p comes to the same value whether the pairs chosen are (B, C) and (A, C) instead of (A, B), so

all is well. Where other treatments exist outside the proposed set, they should be taken into account as well. Thus, in this example:

I	A	A	B	C	D
II	A	B	B	C	D
III	A	B	C	C	E
IV	A	B	C	D	E

it might appear that A, B and C are equivalent. As long as the only pairs considered are (B, C), (A, C) and (A, B) the illusion is preserved because each gives p equal to $\frac{6}{5}$, but when (A, D), (B, D) and (C, D) are taken into account and give values of p respectively equal to $\frac{4}{5}$, $\frac{4}{5}$ and $\frac{3}{5}$, it becomes apparent that C is not equivalent to A and B.

The analysis of data from a totally balanced design is not difficult, but some computer programs that will cope with Balanced Incomplete Blocks break down when presented with the wider class, so care is needed. The full method has been set out elsewhere (Pearce, 1965a). The method of Appendix III can also be used.

32. Some Other Kinds of Balance

Three other kinds of non-orthogonality merit attention. One is supplemented balance, which gives rise to designs described as being of Type S (Hoblyn, Pearce and Freeman, 1954; Pearce, 1960b, 1963). One treatment, usually the control, is in a different position from the rest, which are all equivalent. This facility of giving a special status to one treatment is useful. For example, an experiment may be needed to compare a number of new spray substances with a standard to see if any of them are better, or an unsprayed treatment may be included to establish the existence of the pest and to provide a baseline for measuring efficacy. In the former case it could well be desirable to have more plots of the special treatment than of the others, where in the latter fewer plots might suffice.

An example of supplemented balance is afforded by this design:

I	B	C	D	S
II	A	C	D	S
III	A	B	D	S
IV	A	B	C	S

Here treatments A, B, C and D all have the same replication. Among them any pair gives a value of p equal to $\frac{1}{2}$, but if S is included in the pair the value, p_s say, becomes $\frac{3}{4}$, so A, B, C and D are in fact equivalent. To find the effective replication for comparisons between them, each should be rated p and then p_s should be added to represent treatment S, the whole coming to $4p+p_s = 2\frac{3}{4}$, which is quite good in comparison with the actual replication of 3. In fact, the variance of the difference of two of their means, adjusted for non-orthogonality as necessary, is

$$\left[\frac{2}{2\frac{3}{4}} \text{ (error variance)}\right] \text{ instead of } \left[\frac{2}{3} \text{ (error variance)}\right]$$

as for an orthogonal design. To consider now the effective replication for the supplementing treatment in comparison with one of the others, the figure just found is multiplied by p_s and divided by p, i.e. $4\frac{1}{8}$. That may seem surprising when the actual replication is 4, but the figure relates only to comparisons with other treatments, i.e. the variance of a difference of means of two treatments, one of which is S, is

$$\left[\left(\frac{1}{4\frac{1}{8}}+\frac{1}{2\frac{3}{4}}\right) \text{ (error variance)}\right] \text{ instead of } \left[\left(\frac{1}{4}+\frac{1}{3}\right) \text{ (error variance)}\right]$$

It is in fact nearly as good as for an orthogonal design with the same replication. Clearly the scheme has its advantages. The methods of adjusting treatment means and for analysing data have been set out elsewhere (Pearce, 1965a), though the method of Appendix III can well be used.

In this particular example the supplementing treatment is disposed orthogonally to blocks but that is not necessarily so. One great advantage of supplemented balance is the ease with which awkward numbers can be accommodated. Thus, here there are 16 plots and five treatments so complete equivalence of status is not possible, but the method of fitting everything in is no makeshift. Indeed, by giving more replication where it is needed, the important comparisons have been improved. If equivalence had been insisted upon there could only have been a uniform replication of 3 and even an orthogonal design could have given no better comparison between S and one of the other treatments that would derive from a variance of

$$\frac{2}{3} \text{ (error mean square)}$$

which is not as good as that attained.

To take the second case, there could be two treatments that deserve special status, the commonest example being that of a spraying trial where there is a standard treatment, S, which requires more plots than the rest, and an unsprayed control, O, that can do with fewer. A useful device is to design the experiment in randomized blocks with b blocks and t treatments and to substitute, i.e. to choose at random a plots allocated to O and to apply S instead.

The resulting design is of the kind described as multipartite of Type M (Pearce, 1963). However, it is really simpler than that because two of the parts contain one treatment each, S and O, while the rest is made up of treatments, all of which are orthogonal to blocks. The design has its own special form of analysis (Pearce, 1948), though the general method of Appendix III may be used.

First it will help to work out

$$d = bt + (t-2)a$$

Then the variance of the difference between the mean of S, which will require adjustment, and the mean of one of the orthogonal treatments, which will not, is

$$\frac{2bt + (t-3)a}{bd} \text{ (error variance)}$$

while the corresponding variance for the comparison of an ordinary treatment with the control is

$$\frac{2b^2t + (t-3)a(b-a)}{b(b-a)d} \text{ (error variance)}$$

which are not usually very different from the variances that would be obtained from an orthogonal design with the same replications, supposing that were possible.

A third form of controlled non-orthogonality is shown by the group-divisible designs, i.e. those of the type G (Pearce, 1963). The characteristic is the way in which the treatments fall into m groups each of n treatments, such that all comparisons within groups are made equally well and so are all comparisons between treatments in different groups. To take an example, let an experiment be designed thus:

I	A	B	C	D
II	A	B	E	F
III	C	D	E	F

Here the groups are (AB), (CD), (EF). The two treatments within a group are proportionate and are therefore compared with full efficiency as for an orthogonal design, but comparisons like A *v.* C or A *v.* E clearly need adjustment. To study the design more formally, for any pair of treatments within the same group, e.g. A and B, $p = \frac{1}{2}$, but for any two in different groups, e.g. A and C, $p' = \frac{1}{4}$. The effective replication for comparisons within groups is $np+(m-1)np'$, which in this instance, where $n = 2$ and $m = 3$, comes to 2, the same as the actual replication because of the proportionality property in this particular design. To consider now groups as a whole, let A and B be merged to form X, C and D to form Y and E and F to form Z, then the design becomes

I	X	X	Y	Y
II	X	X	Z	Z
III	Y	Y	Z	Z

i.e. the groups as a whole are totally balanced and are compared with an effective replication of 3 compared with an actual replication of 4.

In brief, the treatments within a group are equivalent; the groups, when the treatments of each are merged, are likewise equivalent. Such designs are useful when the treatments themselves fall into natural groups, as does sometimes happen, especially in factorial designs. For example, in the design above if the treatments A–F are regarded as a factorial set, P1, P2, Q1, Q2, R1, R2, the main effect of numbers and the interaction are studied with full efficiency, but the main effect of letters has an efficiency of only $\frac{3}{4}$. If it were practicable to form two blocks, each of six plots, an orthogonal design would be possible, but where there are restrictions on block size there has to be non-orthogonality somewhere and that means loss of efficiency in some comparisons. It is important to lose the information where it can best be spared.

Group-divisible designs are of many kinds. In some the comparisons within groups are better than the rest and in others they are not as good; some exhibit proportionality where others do not. A general method of analysing data has been given elsewhere (Pearce, 1965a), but the general method of Appendix III is not difficult. Their usefulness has been considered by Kramer and Bradley (1957) and by Bradley, Walpole and Kramer (1960).

33. Row-and-column Designs

In Section 22 it was explained how a Latin square could, as it were, consume its own headland effect by eliminating variation between rows and columns, but it was mentioned also that Latin squares themselves are not often used for this purpose because of the restriction that rows, columns and treatments must all be equal in number. Now that non-orthogonal designs are being considered, it is pertinent to ask if any of them have this same property.

The problem is, as a matter of fact, closely allied to the changing of treatments in a randomized block design. In both, the investigator finds himself with a two-way orthogonal classification, whether of rows and columns or of blocks and original treatments: in both, he wants to add a third classification so as to give optimum precision to the comparison of the new classes. As with the Latin squares (see Section 23), it is assumed that the new set of treatments is unlikely to interact with those originally applied. It may be a matter for surprise that randomized blocks, which are the simpler design, are so much more complicated to change over to a new investigation than are Latin squares, but it should be remembered that the rows, columns and treatments of a Latin square are necessarily all equal in number so the possibilities are limited, but it is only coincidence if the blocks are as numerous as the treatments in a randomized block design. Where this does happen to be true, it is quite easy to effect a change, namely, by regarding blocks as rows, original treatments as columns and using a Latin square.

One possibility is to use a Youden square (Youden, 1937). In this design the treatments are disposed orthogonally to rows but form a balanced incomplete block relative to the

G	F	B	E	D	A	C
C	A	G	B	E	F	D
A	D	C	G	F	E	B
E	G	F	D	C	B	A

FIGURE XIII. Example of a Youden square.

columns (or *vice versa*; there is no real distinction between rows and columns; it's all a matter of viewpoint). An example of a Youden square is shown in Figure XIII, which is only Figure XII modified to ensure that each treatment occurs once in each row. The design would serve equally well for seven new treatments applied to seven blocks each containing four old treatments. It is just a matter of writing blocks for columns and old treatments for rows. It would serve no less well if there had been four blocks each of seven old treatments.

There is no necessity for the non-orthogonal component to be in balanced incomplete blocks. It could, for example, be totally balanced in some more general way as in Figure XIV or group-divisible as in Figure XV. There is a convenient notation that can usefully be introduced at this point. At their first mention non-orthogonal designs have all been given type letters. If O is added to indicate orthogonality, block designs are readily classified, a design in randomized blocks being of type O, one in balanced incomplete blocks being of

A	C	B
C	A	B
B	A	C
C	B	A

FIGURE XIV. Example of a design of type O : OT (i.e. the treatments are disposed orthogonally to rows and in total balance with respect to columns).

B	D	C	A	F	E
A	C	E	D	B	F
F	A	D	B	E	C
E	B	F	C	A	D

FIGURE XV. Example of a design of type O : OG (i.e. the treatments are disposed orthogonally to rows and in a group-divisible design with respect to columns).

type T and so on. Coming to row-and-column designs three letters are needed. The first, which stands before a colon, shows how the columns are disposed relative to blocks and for the purposes of the present publication will always be O. The two letters after the colon show how the treatments are disposed relative to the rows and columns respectively. Thus, a Latin square is of type O : OO, a Youden square of type O : OT or O : TO, the design in Figure XV of type O : OG and so on.

Where a design is of type O : OX or O : XO, X representing any recognizable form of non-orthogonality, there is no difficulty about analysis of data. There must be one classification, rows in the case of O : OX and columns in the case of O : XO, that is orthogonal to the others. For a start it should be ignored, leaving a reduced design of type X, that being the way in which treatments are disposed relative to the remaining blocking system. The data are analysed as if they came from the reduced design, e.g. data from a Youden square as if they came from balanced incomplete blocks. At the end the classification that has been ignored is reintroduced. A sum of squares is worked out for it with the usual degrees of freedom, the method of calculation being the same as if everything were orthogonal, i.e. that given in Section 11. Then this sum of squares and its degrees of freedom are subtracted from the corresponding values in the error line of the analysis of the reduced design to give the error line for the original design. Efficiencies, etc. are the same as for the reduced design of type X.

Designs in which the treatments are non-orthogonal to both rows and columns are more difficult and are in general better avoided unless awkwardness of numbers compels their use. A useful possibility is the design of type O : TT studied by Pearce and Taylor (1948) and shown in Figure XVI. Where treatments are in supplemented balance relative to rows they usually have to be similarly disposed to columns on account of the unequal replication. In fact, designs of type O : SS are readily found, but perhaps the most useful are still those shown in Figure XVII, all of which are formed from Latin squares. Of the three, (a) is a Latin square with a row and a column missing, in (b) a row and a column have been added and in (c) a row is missing but a column has been added. A method of analysis has been given elsewhere (Yates, 1936a) as have methods for (b) and (c) (Pearce, 1952c).

Row-and-column designs provide a fascinating study for those interested in design problems and they are also extremely useful. Where two blocking systems are called for,

B	A	C	D	B	A
D	A	B	A	C	C
A	C	D	B	C	D
C	B	A	A	D	D
C	C	B	B	D	A
B	D	D	C	A	B

FIGURE XVI. Example of a design of type O : TT (i.e. the treatments are disposed in total balance with respect to both rows and columns).

F	E	C	B	D
E	C	A	D	F
C	F	B	A	E
D	A	E	F	B
B	D	F	C	A

FIGURE XVII. Example of a design of type O : SS (i.e. the treatments are disposed in supplemented balance with respect to both rows and columns).

they are often already determined when design takes place and they must be accepted however inconvenient they prove to be, so an awareness of the possibilities is highly desirable. Another paper (Pearce, 1968) attempts to deal with Latin squares as Section 32 dealt with randomized blocks. Some general results, which could well be useful in devising such designs in particular cases, have recently been given (Pearce, 1975).

34. Change-over Designs

With some treatments, such as manuring and pruning, the residual effects can be as important as the intermediate effects and some device for studying them is very desirable. When writing the first edition it was odd to have to record that little attention appeared to have been given to the subject by horticulturists, though in other fields of research, e.g. animal nutrition studies, special designs had been evolved. Now, twenty years after, those same designs appear to be as neglected as ever. Perhaps the methods set out in the last Section have provided a solution. If there are three classifications, (a) blocks, (b) former treatments and (c) current treatments, the last two being unlikely to interact, the natural thing is to study (c) regarding (a) and (b) as blocking systems, but if it is also relevant to study (b) regarding (a) and (c) as the blocking systems there is no objection in principle. Nevertheless, despite their neglect in horticulture and agronomy, change-over designs have been found useful in related fields and therefore deserve mention.

Basically the idea is to keep changing the treatment received by each unit, i.e. the individual animal, tree or other plot, and to do so in such a way that at the end of the experiment each treatment has occurred equally often and has followed every other treatment an equal number of times. A treatment never follows itself. For example, let there be four treatments, A–D, to be applied to twelve units over four periods; the experiment might proceed thus:

Period 1	A	B	C	D:	A	B	C	D:	A	B	C	D
2	B	A	D	C:	C	D	A	B:	D	C	B	A
3	C	D	A	B:	D	C	B	A:	B	A	D	C
4	D	C	B	A:	B	A	D	C:	C	D	A	B

The first period is used for setting up the material: the remainder constitute the experiment proper. In it each treatment occurs three times in each period, once in succession to each of the others.

This particular design is based on three orthogonal Latin squares (Fisher and Yates, 1938, Table XVI) but there are other kinds. Another way of generating change-over designs is to think of the treatments as arranged in a circle, say, ABCD with D as close to A as it is to C, and then to go round following the rule, one forward, two back, three forward and so on. If, now, there are only four units with initial treatments, A, B, C and D, application of the rule to each will give:

Period 1	A	B	C	D
2	B	C	D	A
3	D	A	B	C
4	C	D	A	B

Here each treatment occurs three times in Periods 2–4, and occurs once after each other treatment, but the successions are not equalized over periods as before. However, that is not necessarily a great weakness. If there is an odd number of treatments, it is necessary to have two units initially receiving each treatment. Then in each pair one unit progresses in the usual way, but the other in the reverse direction, i.e. one back, two forward, three back and so on.

Change-over designs have the advantage of using few units. Also, because several treatments are applied to each unit, better comparisons are possible than in the case where each treatment is applied only to certain units. On the other hand, randomization between periods is obviously impossible and the analysis of data suffers in consequence.

Other designs exist, but all are of the same general type as those discussed. For methods of analysis other publications should be studied (e.g. Lucas, 1951; Patterson, 1950, 1951, 1952). There is a useful chapter in the book by Federer (1955).

35. Using parts of trees as plots

In the last Section the problem considered was that of treatments that go on after their application has ceased; in this one the problem is of treatments that have an effect outside the plots to which they are applied. As long as plots consist of whole trees or groups of bushes it should not arise because guards should prevent it, but there are times when an experimenter wants to take as his unit some part of the tree, not the whole of it. For example, he may want to pollinate some blossom trusses with this variety and some with that, or he may want to cut back some shoots but not others. The units will have to be called plots because that is the usual statistical term, though they are far from being plots in any ordinary sense. Also in this Section, "trees" and "blocks" will be used as synonyms. So long as the plots so formed are independent and a treatment applied to one is unlikely to affect other plots, there is no difficulty of a statistical nature. Often, however, there may be fears that a treatment will have remote effects on other plots on the same tree as well as a local effect where it is applied.

The decision whether or not to allow for remote effects is a botanical one. The statistician may have comments to make on the basis of past experience, but the decision must rest fairly and squarely on the experimenter. If he thinks that remote effects are likely he should be most cautious. However, even if he has to recognize that they may exist, there remains the possibility of conducting a soundly based experiment if it is reasonable to assume that the remote effects emanating from any plot will be the same on all the other plots of the same tree and will depend only on the treatment. Again the decision has to be made on botanical grounds and obviously it will depend on the way in which the tree is to be divided into plots. Thus, if it is to be divided into three segments, each representing an arc of 120°, the fact that each proposed plot adjoins both the others might give hope that its remote effects would affect both equally, but if it were proposed to form four quadrants the decision might not be so easy, because then any plot would be in a different position relative to the one opposite than to its neighbours. Again the decision is a botanical one that needs to be made with caution.

If, however, it is reasonable to postulate a local effect and constant remote effects, it becomes possible to conduct a sound experiment. As has been explained, a design affords estimates of treatment effects by comparing plots within blocks. Where each block has its own set of treatments, it also affords estimates by comparing the blocks. In Sections 31 and 32 all the emphasis was placed on the intra-block comparisons, but the inter-block exist as well as was explained in Section 20. In the present context they are indispensable. Let the local effect of a treatment be L and let its remote effect on other plots of the same tree be R, then the intra-block estimate of the treatment effect, W, will give $L-R$, whereas the inter-block estimate, V, will give $L+(k-1)R$, where k is the number of plots in the tree. Hence L may be estimated by $[V+(k-1)W]/k$ and R by $(V-W)/k$. Since the less satisfactory estimate, namely V, carries more weight in R than in L, the remote effect will be less well determined than the local.

In practice it is best to proceed thus: Let there be t treatments and let the plots be such that there are k to a tree. It is now necessary to devise a balanced incomplete block design with b blocks and to use it n times, so that nb trees are needed in all, each tree being divided into k plots. Estimates within blocks cause no difficulty, being made as for any other experiment in balanced incomplete blocks, nor are the inter-block estimates difficult to find, each being equal to

$$\frac{(t-1)}{nbk(t-k)}[(t-k)G-tU]$$

where G is the grand total of data and U the total of data from trees that have not received the treatment in question on any of its plots. From the two estimates it is easy to find the

local and remote effects. Methods of finding standard errors and coping with interactions are presented elsewhere (Pearce, 1957). Goodchild (1971) extended them to cover covariance adjustments (see Appendix IV). Freeman (1973) has extended the range of available designs.

36. Confounding

The subject of confounding is a vast one and scarcely needs description here, when there are so many accounts of it in the textbooks, most of them completely adequate and some truly excellent. The basis of most work is the classical text of Yates (1937), still highly relevant, but other descriptions abound and the best course here, as with other generally understood matters like randomized blocks and Latin squares, is to define briefly what is under discussion and then to discuss it.

Briefly, the idea is to identify certain contrasts that are of little interest and to sacrifice information about them in order to learn more about other contrasts accounted to be of more importance. To take a very simple case, suppose that there are four treatments arising from a 2×2 factorial system, i.e. O, A, B, AB, and, as before, let an asterisk before the designation of a treatment indicate the mean ascribed to it, adjusted perhaps in all sorts of ways, but for the moment that does not matter. Suppose also that someone declares that the interaction of A and B is of no importance; anyone can see that the two factors will act independently. In fact he is declaring as an article of faith that the effect of B will be the same whether A is present or not, i.e. that

$$*B - *O = *AB - *A$$

He might just as well have said that

$$*A - *O = *AB - *B$$

Algebraically the one statement implies the other. For present purposes it will be more convenient to employ a third form and write

$$*O - *A - *B + *AB = 0$$

It is even simpler to write $(1, -1, -1, 1)$, the successive values being the coefficients of the means when taken in standard order.

At this point someone may suggest that the experiment could well be designed in blocks of two plots only, not of four, and he might demonstrate his point, thus: If half the blocks contain only treatments A and B, they give an estimate of $(*A - *B)$ that is unaffected by the blocks chosen. They may be good or bad ones, but that will not affect the difference between the two treatments. Similarly the other blocks can provide an estimate of $(*AB - *O)$. Hence, both the main effect of A, i.e.

$$-*O + *A - *B + *AB$$

and that of B, i.e.

$$-*O - *A + *B + *AB$$

can be obtained, block differences being irrelevant. They are, however, involved in the interaction,

$$*O - *A - *B + *AB$$

since one set of blocks has taken the treatments with the positive coefficients and the other has taken those with the negative and there is no cancellation of block effects. Accordingly, if there were any notable advantage in using blocks with only two plots, the experimenter could avail himself of it, always supposing that he were willing to lose information about the interaction.

However, caution is needed. If no one wants to know about the interaction because everyone agrees that it does not exist, all may be well. If, however, everyone agrees that it

must exist and does not require study because its form is predictable, the design is distinctly faulty. In that case attention will concentrate on the particular effects, i.e. (*A−*O), (*AB−*B), (*B−*O) and (*AB−*A). In each case the two means being contrasted are based on different sets of blocks so it is difficult to think of a design worse for the purpose in hand.

The point needs to be borne in mind when considering the "recovery" of an interaction. As has been seen, one set of blocks contains only treatments O and AB, so the block totals average (*O+*AB), the totals of the other blocks averaging (*A+*B). Hence, if a difference can be established between the two kinds of block, it shows that there is in fact an interaction. Supposing that the pairs of treatments were allocated to the blocks at random, that is easily done. The blocks are regarded as the plots of a completely randomized design with two treatments (A, B) and (O, AB) and a simple analysis is carried out. If there is in fact an interaction, when none was expected, the situation can be serious. Recovery of the interaction can serve as a demonstration that everything is all right, but if everything turns out to have been all wrong the recovery provides no remedy. It is most important therefore to confound only interactions that really are most unlikely to exist.

The situation is different when a main effect is confounded. If half the blocks have only O and A while the other half have only B and AB, the result is a split-plot design. If now recovery, i.e. the main-plot analysis, shows that there is indeed a main effect of B, it is still necessary to look at the main effect of A and the interaction, so no harm has been done. Of course, if there is an interaction it becomes necessary to look at particular effects. Two, namely, (*A−*O) and (*AB−*B), are found within blocks, but the other two, (*B−*O) and (*AB−*A), are effectively lost as before. However, it is to be assumed that the designer foresaw the dangers and was not much interested in them.

An effect is said to be "totally confounded" if all the information about it lies between blocks. However, less extreme cases arise, as in non-orthogonal designs. Thus, for the incomplete block design shown in Figure IX it was noted that the efficiency factor is 0·875 or $\frac{7}{8}$, meaning that $\frac{7}{8}$ of the information lies within blocks and $\frac{1}{8}$ between blocks. In this instance it cannot be recovered because there are six degrees of freedom between seven blocks, which correspond to the six degrees of freedom between treatments, leaving none for error, a state of affairs that is fairly common, so it provides an important exception to the general rule that information lost within blocks can be recovered by studying the differences between blocks. The other extreme case is that of an orthogonal design in which the efficiency factor for all contrasts is one. In it all blocks are made up in the same way with regard to treatments and nothing is to be learnt by studying differences between their means, i.e. all information is within blocks and none is to be found between them. Contrasts are then said to have "full efficiency".

To return to the example of the treatments, O, A, B and AB, three effects were distinguished:

Main effect of A	$(-1, 1, -1, 1)$
Main effect of B	$(-1, -1, 1, 1)$
Interaction of A and B	$(1, -1, -1, 1)$

The interaction can always be found readily because each of its coefficients is the product of the corresponding coefficients of the two main effects, i.e.

$$1 = (-1)\times(-1), -1 = 1\times(-1), -1 = (-1)\times 1, 1 = 1\times 1$$

Similarly, the main effect of A is the interaction of B and A × B, while the main effect of B is the interaction of A and A × B, i.e. if a factor occurs on both sides of an interaction it cancels. Further, it may be noted that if two effects are to be estimated independently, it is necessary for the sum of products of their respective coefficients to add to zero. Thus, for the main effects of A and B

$$[(-1)\times(-1)]+[1\times(-1)]+[(-1)\times 1]+[1\times 1]=0$$

and so for all other pairs. Unless all pairs of effects are independent, the sums of squared deviations for the various effects will not add up to that for the treatments as a whole even though the degrees of freedom do so.

The above relations apply when all treatments are equally replicated and may apply in other cases though caution is required. However, it has been pointed out (Pearce, 1971) that in the totally confounded case two separate contrasts or sets of contrasts are being estimated, i.e. in the example above (*AB−*O) is being studied in one set of blocks and (*A−*B) in the others. There is no necessity for the two parts to be studied in the same way so long as conditions are broadly similar. All that is really required is a trial to estimate (*AB−*O) and another for (*A−*B), then the sum of the two estimates will give the main effect of A and the difference that of B. It is not necessary for them to be obtained from trials of similar design or of the same replication.

In more complicated situations, confounding is still possible and indeed is of maximum usefulness when treatment-combinations are so numerous as to call for blocks that are unreasonably large and there are many high-order interactions that can be discarded. The simplest case arises when all factors are at two levels. Thus, if it were a question of using 2^4 treatment-combinations, O, A, B, AB, C, AC, BC, ABC, D, AD, BD, ABD, CD, ACD, BCD, ABCD, and if the interaction of $A\times B\times D$ can be ignored, it is possible to confound it by first noting that it can be written

$$(1, -1, -1, 1, 1, -1, -1, 1, -1, 1, 1, -1, -1, 1, 1, -1)$$

The coefficients are derived either by writing down those for A, B and D and deriving successively those for $A\times B$ and $A\times B\times D$ or, which comes to the same thing, writing 1 when the number of factors involved (i.e. A, B and D) is even, as in the case of ABC, and writing -1 when it is odd, as in the case of ABCD. If, now, blocks of eight plots are formed and half are assigned the treatment-combinations with the positive coefficient and half receive those with the negative coefficient, the desired interaction will have been totally confounded, all other effects being found with full efficiency.

It is possible to confound more than one effect. In the case of the 2^4 factorial it might be desired to have four kinds of block, *i*, *ii*, *iii* and *iv*, each of four plots, and to confound $A\times B\times D$ between *i* and *ii* as compared with *iii* and *iv* and to confound $B\times C\times D$ between *i* and *iii* as compared with *ii* and *iv*. It speedily appears that the division must go like this:

$A\times B\times D$		$A\times C\times D$	
i, ii	*iii, iv*	*i, iii*	*ii, iv*
O	A	O	B
AB	B	A	AB
C	AC	BC	C
ABC	BC	ABC	AC
AD	D	BD	D
BD	ABD	ABD	AD
ACD	CD	CD	BCD
BCD	ABCD	ACD	ABCD

which leads to

i	*ii*	*iii*	*iv*
O	AB	A	B
ABC	C	BC	AC
BD	AD	ABD	D
ACD	BCD	CD	ABCD

However, there is a third possible way of grouping the kinds of block, namely, *i* and *iv* as compared with *ii* and *iii*, and this confounds $A \times C = (A \times B \times D) \times (B \times C \times D)$, B and D cancelling. This unwanted confounding can be awkward and explains why multiple confounding usually takes place only when factors are numerous and there are many unwanted interactions to choose from.

So far, attention has been concentrated on the case where all factors are at two levels, but there is no necessary restriction. If two factors, A and B, are each at three levels (0, 1 and 2), writing the treatment-combinations in a square thus:

	B_0	B_1	B_2
A_0	00	01	02
A_1	10	11	12
A_2	20	21	22

it is clear that the rows, i.e. (00, 01, 02) *v.* (10, 11, 12) *v.* (20, 21, 22), represent the main effect of A and the columns, i.e. (00, 10, 20) *v.* (01, 11, 21) *v.* (02, 12, 22), represent that of B, each with two degrees of freedom among three groups. The diagonals similarly give the interaction, which has two components, *I* and *J*, each with two degrees of freedom, found thus:

I (00, 11, 22) *v.* (10, 21, 02) *v.* (20, 01, 12)
J (00, 12, 21) *v.* (01, 10, 22) *v.* (02, 11, 20)

It is therefore possible to take three blocks, each of three plots, and to assign them respectively the treatments (00, 11, 22), (10, 21, 02) and (20, 01, 12), thus confounding *I*. If there are a further three blocks available, two courses of action are available, neither very satisfactory. If *I* is confounded again the experimenter is left with *J*, which is not much use by itself, except possibly as something to be tested to confirm that there is no sign of an interaction, at least for those two degrees of freedom. On the other hand, if *J* is confounded in the second set of three blocks, i.e. they are assigned the treatments (00, 12, 20), (01, 10, 22), (02, 11, 20), all degrees of freedom for the interaction are partially confounded with an efficiency factor of $\frac{1}{2}$ and the whole is preserved but not so that it can be effectively studied. Incidentally, a partially confounded design is more usually and more conveniently considered as non-orthogonal and so it is here. If *I* is confounded in half the blocks and *J* in the other half, the data can readily be dealt with using the method of Appendix III.

Given a 3^3 set of treatments in which it is required to confound the three-factor interaction, it is best to consider first the *I* and *J* components of A and B and to write tables thus:

	C_0	C_1	C_2
$I_0 = A_0B_0 + A_1B_1 + A_2B_2$	×	×	×
$I_1 = A_1B_0 + A_2B_1 + A_0B_2$	×	×	×
$I_2 = A_2B_0 + A_0B_1 + A_1B_2$	×	×	×

	C_0	C_1	C_2
$J_0 = A_0B_0 + A_1B_2 + A_2B_1$	×	×	×
$J_1 = A_0B_1 + A_1B_0 + A_2B_2$	×	×	×
$J_2 = A_0B_2 + A_1B_1 + A_2B_0$	×	×	×

Then the *W*-component of the three-factor interaction is formed from the *I* and *C* table in the manner used before for *I* and results therefore in

(000, 110, 220, 101, 211, 021, 202, 012, 122) *v.*
(001, 111, 221, 102, 212, 022, 200, 010, 120) *v.*
(002, 112, 222, 100, 210, 020, 201, 011, 121)

giving three blocks of nine plots. Similarly the *X*-component is formed from the *I* and *C* table using the method for *J*, while the *Y*-component and the *Z*-component are formed from the table for *J* and *C* using respectively the methods for *I* and *J*. Each has two degrees of freedom, making eight in all. Any component or all can be confounded.

Given four levels of a factor it is usually easiest to think of it as if it were two factors each at two levels. As long as the two main effects and the interaction retain equal status, both alone and in their interactions with other factors, in effect the factor with four levels retains its identity. In general, mixed factorial situations, e.g. some factors at two levels and some at three, are very awkward, the one exception being that in which the levels number either two or four, because then the whole can be regarded as if there were more factors, all of them at two levels, with some restrictions on the interactions that it would be acceptable to confound. However, the whole subject has developed a vast literature (e.g. Yates, 1935; Cochran and Cox, 1950; Federer, 1955) which should be studied by those concerned.

37. Interlaced Blocks

When designing a trial it is often difficult to gauge how large the trees will grow, so some means of thinning is desirable to relieve possible overcrowding. Trials with many trees to a plot present no problem because half the trees can be removed without damage to the design, but with one-tree plots special provision is called for, such as the interlacing of blocks.

This simply means that the tree positions are divided into temporary and permanent, some blocks being confined to one and some to the other, a device that is legitimate where randomization in all blocks is unrestricted, as it should be. No modification of the statistical analysis is called for. Figure XVIII shows a peach variety trial in which this scheme was used. It is not available where plots need to be guarded.

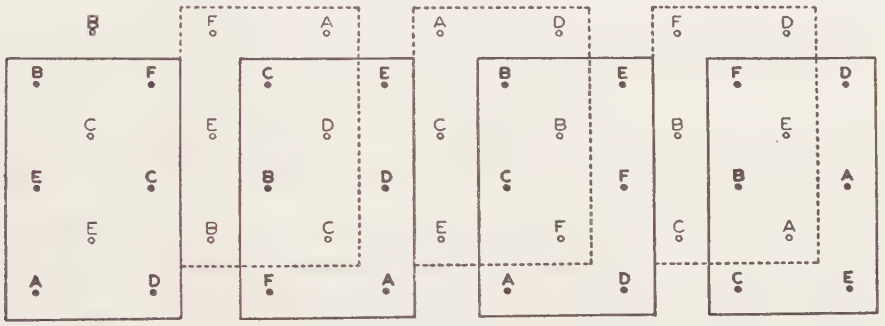

FIGURE XVIII. Example of a trial with permanent and temporary trees in interlaced blocks.

When the number of treatments is a multiple of four it is often a useful device to form them into four groups designated O, A, B and AB and to take the blocks in pairs. Effectively that gives a group-divisible design. On Blocks I and II, say, let the main effect A be confounded between temporary and permanent trees, the treatments applied to the temporary trees of Block I being assigned to the permanent trees of Block II. Similarly, the main effect of B might be confounded on Blocks III and IV and the interaction on V and VI. If there are yet more blocks, some of the effects can be repeated, while if there are fewer, some need not be confounded at all. Now, when the trial is first planted, the main effect of A will be partially confounded between Blocks I P and I T and between Blocks II P and II T. When thinning

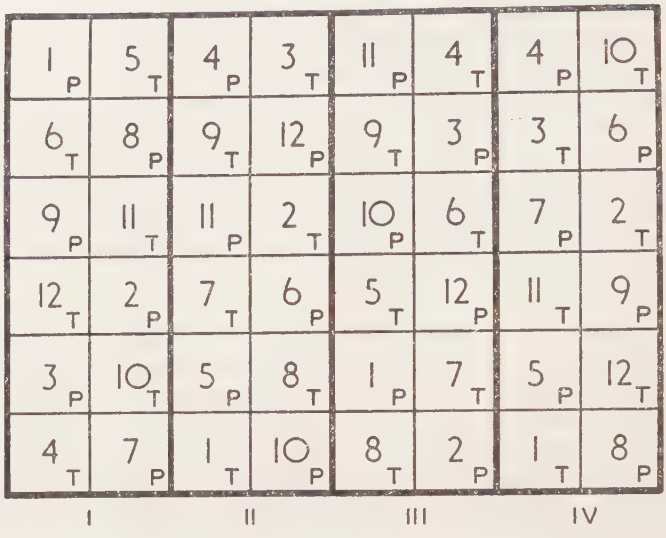

FIGURE XIX. Interlaced blocks with partial confounding to permit thinning of a trial with one-tree plots.

has taken place, the same effect is partially confounded between Blocks I P and II P. An example with twelve treatments is presented in Figure XIX, treatments 1–3 forming Group O; 4–6, A; 7–9, B and 10–12, AB.

The same device is sometimes valuable in raspberry variety trials. On account of the tendency of some varieties to send up suckers, known in England as "spawn", some distance from the parent plant, such trials have to be kept under frequent observation so that contamination of varieties can be prevented; but the task is made much more difficult if the randomization places two varieties with indistinguishable spawn side by side or end to end. To prevent this, it is enough to arrange that whenever two such varieties are to be compared, they shall always be in the same group. Then, by dividing the plots of each block into sub-blocks on the diagonals like the permanent and temporary trees of a plantation and by partially confounding as already described, no pair of indistinguishable varieties can ever occur on plots side by side or end to end. It is necessary, incidentally, to keep fairly wide cultivated alleys (usually across the rows) between the blocks, for otherwise there is nothing to prevent two adjacent plots in *different* blocks giving similar spawn. The whole subject has been discussed in some detail by Taylor (1949). Figure XIX illustrates this use of partial confounding no less than the problem for which it was introduced, the only difference being that P and T now indicate the plots of sub-blocks and not permanent and temporary trees. It is not necessary, however, to confound each effect in two blocks, one being sufficient.

38. Some Disadvantages of Complex Designs

Some writers have recommended the use of more complicated designs to control environmental variation, giving as their reason the great expense incurred anyway in conducting a long-term trial and the consequent need to economize as far as possible in the number of trees to an experiment. The present writer appreciates the point of the argument but believes that the case for complexity is not unassailable.

In the first place, for the reasons given in Section 14, it is by no means certain that smaller blocks do in fact give a worth-while reduction in experimental error with perennial species. Sometimes they do, but not always. The expectation of success depends both upon the magnitude and nature of the environmental effects to be controlled and also upon the extent of the variation inherent in the plant material. Thus, an experimenter working with variable non-clonal material like coconuts or oil palms probably has most of his variation coming from the plants, and even a complete elimination of the relatively small residuum coming from position in the field might benefit him very little.

For another, in a long-term trial mishaps are much more likely than in a trial with annuals. Whole treatments may fail, some of the plots will almost certainly be lost and corrections by covariance are common; hence even designs simple in the first place become complicated, while trials initially complex become ultimately unmanageable. In addition, trees often outlive the investigation for which they were planted and it is a great advantage to be able to use them for a further and quite different investigation.

Nevertheless, more complex designs have their place. Thus, it does sometimes happen that a better control of environmental variation is to be expected from using smaller blocks and the trial is unlikely to last long enough for many mishaps to occur. Figure XII illustrates such a trial. Again after a set of trees has fulfilled its original purpose, a further and more complicated trial is sometimes applied to it, using the methods of Sections 23 or 32. That makes for complexity, but with established trees many of the mishaps will already have occurred, so the investigator may feel that further deaths are unlikely. Decisions of this sort, naturally, depend upon the species and its principal diseases. With some, such as plum and cherry, losses when they occur mostly come soon after planting. Further, with older trees, the environmental effects will have been operating for some time and their elimination is all the more likely to be worth while. In general, complicated designs are of greatest value in short-term trials on established trees, though even there calibration (Chapter 4) may provide a better solution.

39. Nonce-designs

Every experimenter has had his occasional bright idea for conducting an experiment and has thought about using some scheme that certainly did not come out of any textbook. When people use words made up on the spur of the moment to meet an unusual situation, etymologists

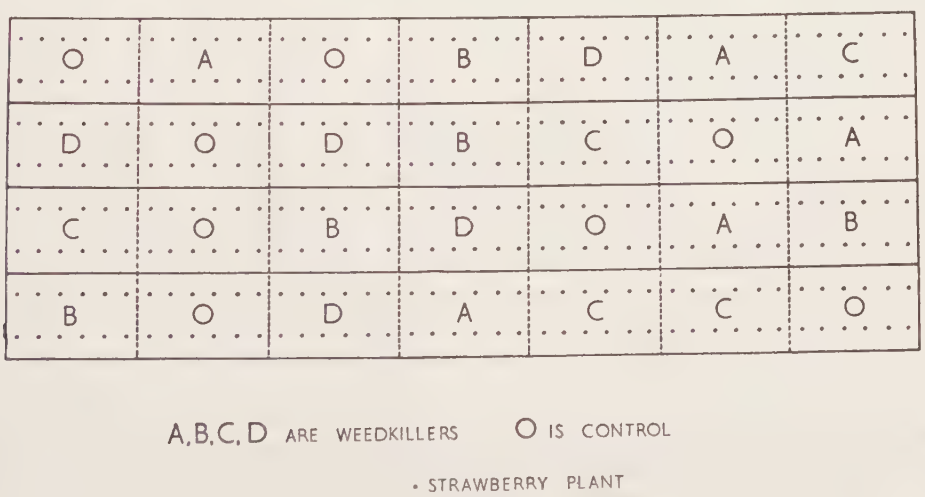

FIGURE XX. Example of a herbicide trial designed in supplemented balance.

speak of "nonce-words"; equally one can speak of "nonce-designs", which are used once for a particular occasion with no idea of their being used again or generalized into anything wider.

Thus, a number of years ago a trial was designed at East Malling in which there were to be four randomized blocks, each of seven plots, for the comparison of five weed-killers against an untreated control, which was to have two plots in each block. At the last moment, when all the plots had been pegged out, one of the weed-killers was held up in transport and the design in Figure XX was substituted. It was in fact the first design in supplemented balance, but that was not foreseen at the time. Incidentally the long narrow blocks were chosen on account of the initial distribution of weeds.

Nevertheless, the invention of new designs is a tricky matter and most nonce-designs prove, on mathematical examination, to have some defect. The commonest faults are undue restriction of randomization leading to bias in the error and an inability to partition the treatment degrees of freedom in any satisfactory manner, but there are others. An experimenter without mathematical training is well advised to keep to designs vouched for by those able to judge. To take an example, a design known as the semi-Latin square† has been "invented" again and again, among others by the present writer, though it is known to give a biased error (Yates, 1933). The fact that a means has been devised for setting limits to the effect of bias (Ma and Harrington, 1949) does not really alter the point.

That is not to say that there are no new designs to be discovered. Indeed, at the moment, a much freer approach to design problems is becoming apparent but experimenters need to exercise great caution with it.

† By a "semi-Latin square" is meant a layout in which there are twice as many columns as rows, each treatment occurring once in each row and once in each pair of columns, e.g.

A	B	C	D	E	F
E	D	A	F	B	C
C	F	E	B	D	A

CHAPTER 4

CALIBRATION

"I do not know whether there is anything peculiarly exciting in the air of this particular part of Hertfordshire, but the number of engagements that go on seems to be considerably above the proper average that statistics have laid down for our guidance."

OSCAR WILDE, *The Importance of Being Earnest*, Act III (Lady Bracknell).

40. The Idea of Calibration

As has already been emphasized, the variation of perennial plants comes from two main sources, position in the field and factors inherent in the plants as individuals. The last two chapters have been concerned with the allocation of treatments to plots and thus with the control of environmental effects, but is there nothing to be done about the second source of variation?

If some trees are inherently better croppers or more vigorous than others, the fact, if it is to be of use, should be ascertained before applying treatments. After the event it is no good trying to explain treatment differences by suggesting that a certain treatment happened to be on better trees, for that is to invade the province of the significance test; the position is different if there was prior reason for expecting one set of trees to do well. Then, if they are undistinguished, the treatment may be judged a poor one. It will be adjudged good only if the trees do excellently. That is to say, the investigator starts his trial with a forecast of what he can expect from his plots and he judges his actual figures in the light of the expected ones. The future performance of trees naturally depends upon future weather, so he cannot say: "Over the next four years I shall expect to get 120 kg of nuts from this plot and 100 kg from the next," but he can often be fairly certain that the first will do better. For example, he can take measurements at an early stage in the development of the plots and judge future performance from them. Such forecasting is called "calibration" (Pearce and Taylor, 1950). The idea is very old (Chandler, 1921; Vaidyanathan, 1934) and has had widespread application.

Part of its value lies in its providing some control over plant variation that has not been eliminated by initial selection. If trees do vary a lot in the nursery, the experimenter's first reaction will be to reject extremes if the purpose of the investigation allows him to do so. His next might be to grade and allocate different grades to different blocks. If there is still appreciable variation he can take records that show it and apply corrections later, i.e. he can calibrate.

However, as has been explained in Section 14, with many species, plants that are initially complete often even up if given long enough, so very early records are often poor calibrators. A further use for the approach, and probably its greatest, comes when trees have already grown for a time and are beginning to show their future potential, i.e. the slow build-up of environmental variation is beginning to dominate other sources. It follows that blocking and calibration are not to be seen as having unrelated uses, the one controlling environmental variation and the other controlling the biological variation of the plant material. Once a tree has become established it has its own set characteristics, which derive in part from the local conditions, in part from itself and in part from accidents of planting, pruning, etc. To some extent blocking, which depends upon spatial patterns, and calibration, which depends upon temporal ones, are rivals because environmental variation has both a spatial and a temporal aspect, but more usually the two approaches are complementary.

Where there has been a change of treatments (see Sections 23 and 32) it is sometimes possible to take useful calibrating measurements just before the new set of treatments is applied. Where the experiment is newly planted, some writers have thought that trees can be calibrated by previously growing some annual crop, such as wheat, on the experimental

area and harvesting each of the future plots separately so as to gauge their individual fertilities. There may be something in the idea, but the present writer has yet to hear of an instance where it has worked.

To calibrate to best advantage the trees should be planted and left for a time without differential treatments, but that is often difficult, if not impossible. Thus, in a variety trial the treatments are necessarily applied not merely at planting, but earlier while the plants are still in the nursery. With ungrafted material, indeed, the differential treatments have always been present. Some treatments, again, have to be applied early if they are to operate at all. Supposing that a preliminary period is possible, care is needed in applying a uniform treatment that will not interfere with the comparison of future differential ones. Thus, if trees are irrigated during the preliminary period their roots may well be nearer the surface than if they had been left unirrigated, which may affect the response to subsequent treatments. For that reason, treatments during the calibrating period need to be chosen carefully so as to produce trees typical of the region and interpretation of the subsequent data must be made in the light of the history of the trees.

It should perhaps be emphasized that calibration is not a cause of delay in securing experimental results but rather a means of saving time. As will appear in Section 42, it is possible to lay out an orchard for calibration in such a way that the types of trial for which it can be used are scarcely restricted at all. It follows that, if a research station can maintain a supply of such calibrated trees, an experiment can be started at once without waiting for suitable trees to be raised and planted, followed by a further delay waiting for them to come into bearing. However, helpful as these designs are, effectively they presuppose that one use of the trees is as likely as any other and so they make preparations for all eventualities. In practice the position is rarely so open. It might, for example, be intended to use the plants for the study of pruning methods, in which case plots will probably be formed of single unguarded trees, or for a fertilizer investigation, in which case plots will be larger to allow of economical guarding, or for a comparison of spray schedules against insect pests, in which case allowance will have to be made for spray drift and the spread of mobile species. Where an intention is thus implicit, there is no need to provide for other eventualities and designs can be simplified accordingly.

The calibrating measurements having been made, the analysis of covariance usually provides the best means of using them (see Appendix IV).

41. Covariance and Some Possible Alternatives

In essence the method of covariance (Wishart, 1950; Cochran, 1957) involves two parallel bodies of data. The one, called the "dependent" variate, represents the character under study; the other, called the "independent" variate, represents some other character that is suspected of disturbing the results of the experiment and which arises from sources irrelevant to its purpose. The method allows the dependent variate to be adjusted by the independent, using a formula derived from the data themselves. Hence, if the independent variate is not in fact causing any disturbance, no adjustment will be made. If it is, the adjustment formula may not be of general application but it will be relevant to the matter in hand, which is the important point.

To take an example, a commercial plantation of established apple trees is to be taken over for experimental purposes and it is recognized that the trees differ greatly in size. It is therefore proposed that the trunk circumferences shall be measured before treatments are applied to provide the independent variate, the crop to be weighed in the future providing dependent ones.

There is no certainty that the trunk circumferences will be related to future crops in any way and, if they are, there could be doubt what the relationship will be. If the trees differ in size on account of depth of soil, it is to be expected that those with thick trunks will be those favoured by their sites and consequently those with higher cropping potential. However, it

could be that the soil is of nearly uniform depth and the trees with thicker trunks are those of more upright habit that have better access to light. Once they have borne fruit and their branches have been brought down, their crop may be large indeed but until that happens cropping may be poor. A sensible experimenter will not dogmatize about the form of the relationship but will leave it to the analysis of covariance to find the adjustment formula, if indeed adjustment is needed at all.

Strictly speaking, calibration is the use of records taken prior to the application of treatments on the plants themselves to form the independent variate in the analysis of covariance, but it is closely related to the use of site characteristics, such as depth of soil or available nitrogen or drainage or the performance of previous plants as independent variates, e.g. Holland (1956) considered soil acidity in relation to growth of apple trees. Such measurements have the advantage that they are not affected by the treatments, so data for them can be recorded while the experiment is in progress. For example, a raspberry variety trial showed marked variation in growth and cropping that, on investigation, appeared to be related to the soil texture. Samples quickly gave an independent variate that removed three-quarters of the error variance and led to an experiment that was rather more precise than most. Of course, that is no way to conduct research. Nevertheless, the belated calibration saved the day as it has done on other occasions.

The price of a covariance adjustment is a good deal of computing labour and the loss of one degree of freedom from the error line of the analysis of variance. In certain circumstances the cost of either item may appear exorbitant but usually the price is reasonable.

There is no need to base adjustments on only one independent variate and sometimes it is advisable to include several. Each additional variate leads to an increase in computing and the loss of a further degree of freedom so there has to be a limit to the number, but there is no need to stop at one. It may be noted that the amount of computing goes up more with the square of the number of variates rather than with the number itself. A further limit may be set by the restriction of computer programs, which often will not deal with more than a few independent variates at a time.

Given a dependent variate it is tempting to try a range of possible calibrating variates and to adopt the one that appears to be most effective, but that can be a dangerous procedure. If enough sets of random numbers were tried, sooner or later one would be found to fit the dependent variate by chance, but its adoption could only mislead. It may also be noted that when two or more calibrating variates are related among themselves they are sometimes more effective as a group than would be expected from their effectiveness as individuals. For example, the trunk circumferences at the start of an experiment may well be correlated with the crops in the previous two years. It may not be clear which will better calibrate crops over the next two years, but a double covariance adjustment on the two may be better than a single one on either alone.

Despite the obvious advantages of the method of covariance there is a reluctance in some quarters to use it, but in general the alternatives are inferior. For example, instead of allowing the data to suggest the adjustment formula, an assumption may be made. Instead of using crops after application of treatments as the dependent variate and previous crops as the independent an experimenter might decide to analyse the ratio of yields in the two periods or perhaps their difference. Even if the two are closely associated, little will have been gained as compared with the use of covariance; if they are only loosely associated, something may have been lost. Indeed, if some trees have had excessive early crops that have reduced growth, a variate will result that is both extremely variable and possibly misleading.

To take some examples, if some strawberry plants in a plot die it is not to be assumed that the crop will be reduced in proportion, because the plants that survive may well be larger. If some trees have more blossom trusses than others it does not follow that they will give fruit in the same proportion. In both these instances it would be safer to use covariance and let the data provide the basis of adjustment.

Another common way of avoiding covariance is to form the blocks of the experiment on the basis of past performance (e.g. Koch and Hunter, 1957). Although not positively wrong it is in general not to be recommended for reasons that will be discussed in Section 50. Obviously it would be unfortunate to form blocks on the basis of a record that proved not to forecast future performance, whereas an ineffective covariance adjustment does little harm.

To decry the alternatives is not to argue that the method of covariance is perfect, but its disadvantages can mostly be avoided by exercising care. However, the method assumes that the regression coefficient, i.e. the effect upon the dependent variate of a unit change in the independent, will be the same whatever the initial value of the independent variate and whatever treatment is applied, but neither assumption is necessarily justified.

To consider the first point, it is unusual for the two variates to be related in a genuine straight-line relationship, but a reasonable approximation is desirable and can often be attained by transforming the independent variate to its square or its square root or its logarithm or some other function of itself to obtain the deserved linearity. Sometimes even that is of doubtful value and it is then a good plan to use two variates, one the original and the other a transformation of it. Thus, Abeywardena (1964), working with coconuts, wished to adjust crop by x, the number of bearing palms per plot, and achieved good results by a double covariance on x and $\sqrt{x}$, which allowed adequately for the curved relationship.

The second point requires more care. To take an example, in general it is reasonable to adjust later crop yields by earlier ones, but in a fertilizer trial it could well be that the relationship between the two is different when there are sufficient nutrients from that when the plants are starved. The point does not often arise, but it can be so serious when it does that it must always be borne in mind.

Although in general the independent variate should either be measured on the plants before the application of treatments or else represent some permanent characteristic of the plots measured whenever it is convenient to do so, certain exceptions can be admitted. Thus, in the coconut example of Abeywardena, given above, the number of palms bearing at any time was considered to be genetically determined and therefore unaffected by the treatments. Even when the treatments may have had an effect on the independent variate, the method of covariance may be justified, though care is needed. Thus fertilizer applications on tea bushes may lead to increased growth as well as increased yield. To adjust the yield by the growth is to enquire whether the fertilizer has led to more crop per unit of growth, which would elucidate a point of some interest. There is no objection to doing so, provided the experimenter is clear where his analysis is leading him. He will then avoid incidents like the one that is said to have raised doubts about the method of covariance when it was first introduced. The story goes that an experimenter adjusted the yields of a barley trial by the germination rates and so reached the conclusion that the differential treatments were having no effect on yield, which was obviously untrue. The correct conclusion should have been that the treatments were having no effect on yield other than that brought about by improved germination, which would have been a useful result.

42. Lay-out of Plantations for Calibration

When laying out a plantation for calibration it sometimes happens that its ultimate use cannot be foreseen. Thus, it may eventually be used for a trial of soil management, which will need large plots with guard trees, or it may be used for a pruning experiment, in which the treatments are likely to be applied to single unguarded trees. Since calibration takes some time, a research station is well advised to keep a stock of trees to hand, previously calibrated, ready to be made use of experimentally; but if this is done the design of these plantations must be as "adaptable" as possible.

A little consideration will show that some numbers are in themselves more adaptable than others for the number of trees along one side of a rectangular trial. Suppose, for example, that a plantation has 285 trees, arranged 19 by 15. Assigning headland trees to form outside

guards, the rest are arranged 17 by 13. Both these numbers being prime, allocation of the trees into blocks and plots is going to prove exceedingly difficult if waste is to be avoided; but had the trees been arranged 20 by 14, to leave 18 by 12 after the exclusion of guards, the possible divisions are numerous. In fact, 20 and 14 are "numbers of high adaptability", whereas 19 and 15 are of "low adaptability". In designing a calibration trial it is often worth going to some lengths to obtain suitable numbers of trees. Thus, given an area 95 metres by 75 and trees usually planted at 5 metres square, the natural thing would be to use the numbers

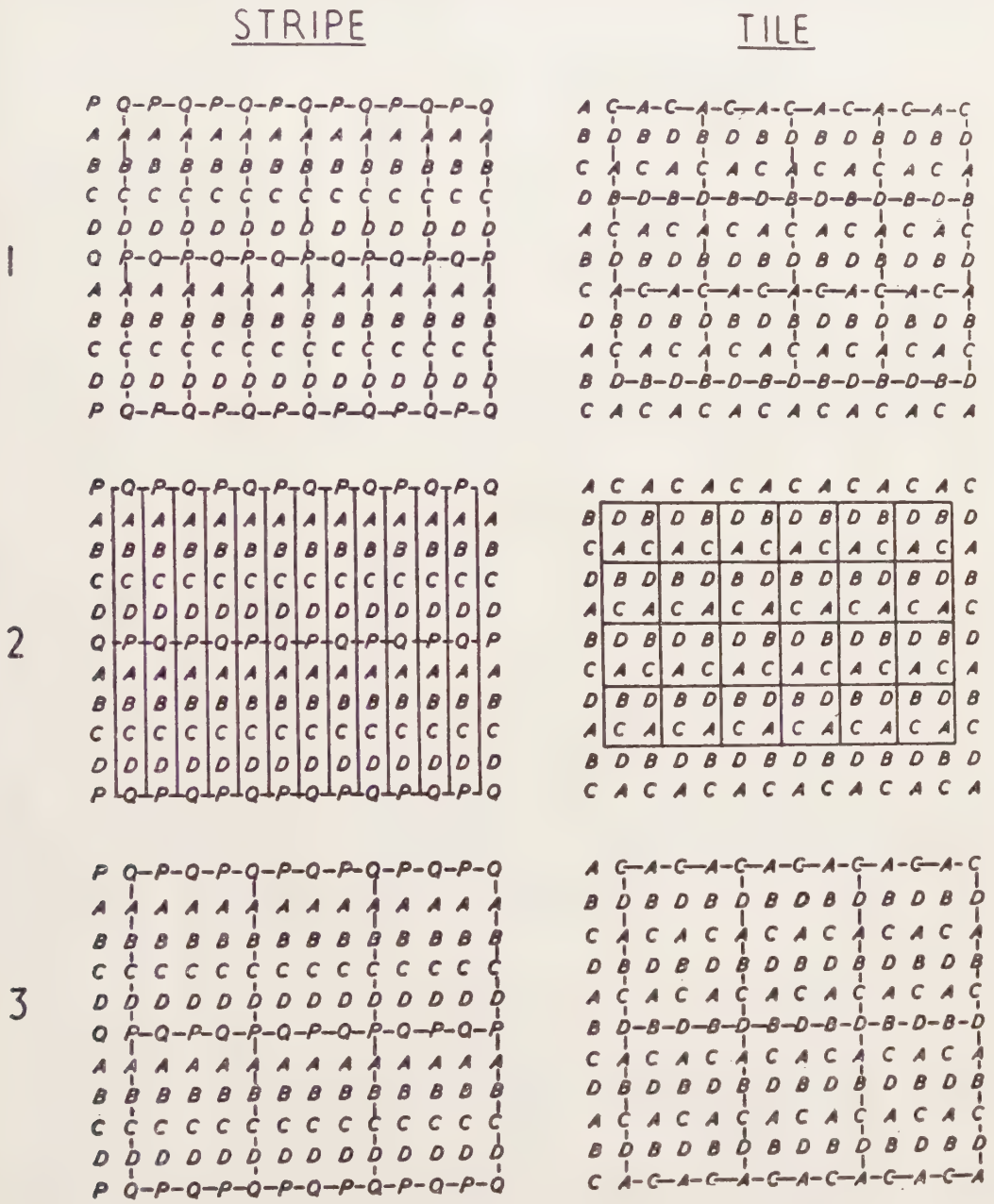

FIGURE XXI. Examples of calibration areas designed in stripes and tiles.

of low adaptability referred to above, but with a little thought it might appear better to plant at 4·75 metres by 5·25, so as to make the dimensions of the plantation correspond to numbers of high adaptability. The whole question has been gone into at greater length elsewhere (Pearce and Taylor, 1950).

In designing a calibration trial it is often advisable to include a range of varieties. With some species it is in any case essential for pollination; but, even where this consideration does not apply, a range does enable subsequent trials to be widely based in their conclusions. The varieties should be selected to represent distinct classes, as different from one another as possible but suitable for growing in the same plantation. Unless there is a strong reason for the contrary, the varieties chosen should be of commercial importance, but sometimes it is not possible to arrange that. Two special types of layout, "stripes" and "tiles", have been suggested for use in such experiments.

Stripe designs have been described elsewhere (Pearce and Hoblyn, 1948; Pearce and Taylor, 1950). The varieties are planted in parallel lines down the larger dimension of the plantation, the lines being grouped into stripes, each stripe containing one row of each variety, the allocation of varieties to rows within a stripe being systematic and not at random. The stripes are separated by rows made up of varieties chosen so as to ensure good pollination of those in the experiment. A typical example is illustrated on the left-hand side of Figure XXI, which illustrates ways of forming plots. Stripe designs have the disadvantage of possibly wasting the pollinator rows, though they can readily be used as guards if any are needed. Stripe designs have the compensating advantage that the main varieties do not have to pollinate one another and may therefore be chosen freely.

In designing trials in stripes, it is important that the rows should contain trees to a number of high adaptability. Also, the number of stripes should, if possible, be a power of two or three or a product of these numbers, such as 2, 3, 4, 6, 8 or 9 rather than a number like 5, 7 or 10. The aim should be to obtain a useful number of plots whether guards are needed or not, and in general the best numbers are those of the form $2^m.3^n$ where neither m nor n is small. To take an extreme case, 35 plots can only be used for 5 replicates of 7 treatments or 7 replicates of 5, while 32 plots can be used only for 2, 4 or 8 treatments in 16, 8 or 4 replications; but 36 are available for 2, 3, 4, 6 or 9 treatments in 18, 12, 9, 6 or 4 replications—a much wider range.

In such designs the varieties will be disposed systematically, but that can be allowed for, as was shown in Section 27.

In tile designs, which have been described in some detail elsewhere (Pearce and Taylor, 1950; Pearce, 1952b) the idea is to start off in one corner with a pattern of varieties, known as a "tile", and to repeat this pattern both along the rows and across them until the whole experimental area is filled. There is no objection to parts of tiles to fill in the edges. The designs are so called from the resemblance to laying a floor with tiles of a uniform pattern. An example is set out in the right-hand side of Figure XXI, from which it will be seen that plots may be formed in several ways, whether guard rows are needed or not. Those shown in the top and middle diagrams are square, but plots four trees long are also possible if arranged vertically.

If there are only two varieties the recommended tile is

A B
B A

It simply places the two varieties on the diagonal.

If there are three varieties, it is possible to adopt a similar tile thus

A B C
C A B
B C A

It leads to a very adaptable design, though it is sometimes better to duplicate one variety and to use this instead

A C B C
B C A C

For four varieties, a good tile is

A B C D
C D A B

the one illustrated in Figure XXI. While for six a possibility is this

A B C D E F
D E F A B C

or this

A B C D E F
C D E F A B
E F A B C D

The design of calibration trials in tiles calls for quite a lot of thought, and the references given will suggest some details of technique for the various types of tile. The basis is always the same, however. The number of trees to the longer headland should be adaptable and the number of plots arising should, if possible, be of the form $2^m.3^n$ whether guards are used or not.

Again the approach of Section 27 can be used to allow for the lack of randomness in the placing of the different varieties.

43. Some Experiences with Calibration

Numerous writers have reported successful applications of the method of calibration and a summary may be useful as well as encouraging. There are, however, others who have tried without success but they have not written papers about it. Nevertheless, with most important species there are indications that could well be followed up.

One species about which much is known is the apple. Sudds and Anthony (1929) calibrated crops by means of trunk measurements and concluded that the area of cross-section was better than the circumference, but Wilcox (1940) found $(\text{circumference})^{1.5}$ to be best. It is true that he was relating crop in a two-year period with trunk measurements, so strictly speaking he was not calibrating, that not being his purpose, but his results find confirmation (Pearce, 1949) in the observation that circumference and area of cross-section, i.e. $(\text{circumference})^2$, were about equally effective, the trees in this instance having been measured beforehand. Butijn and van der Boon (1957), however, found the length of fruit-bearing branches a better calibrator than either the crown diameter or the trunk circumference. Later Pearce and Brown (1960) found a suggestion that trunk records are of greatest use after a calibrating period of poor crops but good growth, whereas previous crops are of greater value after trees have been yielding well. That may relate to the experience of Clarke (1967), who tried calibrating crops of two varieties of apple and found that the trunk circumference at the start of the experiment was better for one, but its rate of increment was better for the other. Returning to the paper of Pearce and Brown, it was recommended as a general procedure that apple crops should be calibrated by double covariance on both past crops and initial trunk circumference. It was also noted that after a plantation has been thinned the behaviour of the remaining trees is difficult to predict for some time afterwards, so calibration is apt to fail just when it is most needed, i.e. when the replication has been reduced and variability

is high on account of disturbance. Working on commercial smallholdings Abraham, Hoobakht and Nejad (1971) found trunk circumference a useful calibrator for crop.

Turning to previous crops as a calibrating variate, Collison and Harlan (1930) found for apples that crop over a five-year period was effectively reduced in variability by taking account of crop in the preceding five years. Yeager and Latimer (1940) also preferred previous crops to trunk measurements. The work of Wilcox (1937a) also repays study for its thorough investigation of such relationships in general.

Although on the whole it is difficult to find early calibrating records for apples, Moore (1962) has pointed out that trees that carry a very early crop and so spread their branches and come into bearing earlier are later overtaken by those that were at first upright and come into full bearing later. He suggested the measurement of branch angles.

There is also evidence (Pearce and Brown, 1960) that cropping patterns with apples do not persist indefinitely. A tree that bears well can inhibit its own growth and so bear less well in future. Consequently, calibrating periods should not be too long and it is easier to calibrate crops over short periods (e.g. two years) than over long (e.g. six years). Grosse (1941) has given an interesting account of the relationship between growth and cropping in successive years.

Nevertheless, in a study of pears (Pearce and Brown, 1960) the cropping pattern proved to be remarkably persistent but, even so, there were times when it could usefully be supplemented by the trunk circumference. Accordingly the same recommendation was used as for apples, namely, to use double covariance.

Another species that has attracted much attention is cocoa. Cheesman and Pound (1932) related the number of pods in a single season to previous crops and concluded that the calibrating period should be as long as possible, commenting that very large crops appear to be as individualistic in distribution as very small ones. Vernon and Morris (1964) make the useful point, which applies to many other species, that when biennial cropping occurs it is better to use periods containing an even number of years. They found a single year's crop to be a valuable calibrator, but they recommended two, provided that does not cause delay. They were working with native small holdings, as were Pearce and Thom (1951b) who found a correlation between crops over successive periods of two or four years but concluded that calibration should not go on too long. They also found useful correlations between crops and the numbers of trees in a plot, the planting being irregular. Cunningham and Burridge (1959) likewise found correlations between crops in successive two-year periods but considered single season crops unsatisfactory. Longworth and Freeman (1963) recommended (a) trunk circumference when there was no time to measure crop and (b) the use of the two variates together as trees became older.

Tea might be expected to exhibit some special problems because the crop is the leaf, not the fruit. In fact, Eden (1935) found useful correlations within a pruning cycle. Sen (1963a), working with young tea, correlated successive year's crops. He superimposed various experimental designs on his uniformity area and it is interesting to note how, after calibration, the error variance was much the same for all designs, though it differed before, i.e. calibration had been complementary to blocking. Later (Sen, 1963b) he reached the same conclusion for plots of varying shapes and sizes, which were more nearly of the same effectiveness after calibration, than before. All records were taken within one pruning cycle. Later, Sen and Biswas (1966) found that calibration of yield ceased to be effective if long experimental periods were used. They considered yield to be a better calibrator than weight of prunings but obtaining it could waste time. Both variates, however, could be effective but sometimes the one and sometimes the other was better.

Another species with an unusual crop is rubber. Taylor (1926) studied annual yields in relation both to past yields and other measurements and found some high correlations. Murray (1947) found that yield in a single season was related to earlier yields if recent. Narayanan (1966) studied the relationships between trunk circumference and yield at different

times and later concluded (Narayanan, 1968) that yield was well calibrated by trunk circumference, but for longer experimental periods it was better to record previous yields and to use double covariance. Later (Narayanan, 1970) he repeated that recommendation, noting that if only one calibrating variate were allowed, sometimes the one and sometimes the other would be better.

With oil palms Chapas (1961) found calibration useful provided the preliminary period did not extend over more than four years. When it was used, the association between adjacent plots was less, a result shown by the higher value of b in Fairfield Smith's expression (Section 14).

Lutz (1939), working with early pecan crops, found a relationship with the cross-sectional area of the trunk at planting. Sharpe and Blackman (1950) worked with pecan and found the crop over the past five years to be better than either trunk circumference or cross-sectional area, which were about equally effective.

For cloves Tidbury (1943) found trunk circumference useful but recommended that it be supplemented by a measure of canopy.

For coconuts Iyer (1958) makes the points that cropping is commonly biennial and studies are aided by calibration. Shrikhande (1958) recommends calibration but argues that two-year periods are needed when there is a biennial tendency. Abeywardena (1962b) suggested that varying numbers of trees to a plot could be dealt with by covariance upon a function of the number of survivors. An anonymous article (Anon., 1961) makes the point that with coconut, mango and arecanut calibration can go on for too long. A later report (Anon., 1962), however, found calibration ineffective with mango. Also for arecanut Agrawal, Bavappa and Khosla (1968) found it possible to calibrate one year's crop by that of the two preceding years.

Butters (1964) found for coffee that crop was calibrated by the stem diameter but questioned whether it needed to be recorded with high precision.

For oranges Parker (1942) concluded that crops over a four-year period were well calibrated by earlier crops, if recent, but not by measurements of size of tree. Parker and Batchelor (1932) had already found that crops in single years were related to earlier crops, but when trees were young their yields gave little indication of future performance.

With other fruiting plants Reed (1929) found that single-year crops of apricots were related to the increment in trunk circumference during the preceding year, while Taylor (1951b) found a useful calibration procedure for strawberry crops by taking a double covariance on heights and spreads. Also, where plants had been deblossomed, he found a correlation between the weight of blossoms removed and the following crop. Similarly, Rigney, Morrow and Lott (1949) recommended the number of shoots of blueberries and the visual ratings in grape vines as calibrators. Goodchild and Sprent (1966) recommended double covariance on the previous year's crop and the number of canes tied in for raspberry crops.

It is unfortunate that well-established recommendations are sparse and rather unrelated but certain general principles appear to find support. For example, it is clear that cropping patterns tend to be erased, so the total length of calibrating period and experimental period together must not be too long. Again, where there are biennial effects neither the on-year nor the off-year crop alone shows much relation to yield in neighbouring years, but two-year periods are much better. Finally, though the point is less well determined, there does seem to be something in the idea that the relationship is between the dominant activities in the two periods. Thus, after a period of growth, tree size sums up the potential of the individual plants unless increment in tree size is better, but after a period of cropping it is the crops that provide the better calibrating variate. In this connection it will be noted that success with plant size as a calibrator commonly appears when trees are young and emerging from their vegetative phase.

Of course, calibration is available when other records than crop are under study. It is a pity that so little is known about the calibration of, say, trunk circumference and and blossom count and much else.

44. Trials on Commercial Plantations

One of the principal uses of calibration comes when designing a trial on a plantation that had originally been laid down commercially. The careful standardization of material and cultural treatments of an experimental plantation has usually been lacking and injured trees may not have been replanted. In such circumstances the prospect may well appear hopeless. There are, however, usually compensating features. For example, there is only rarely a shortage of trees, so the experimenter can concentrate on getting as much information as he can from each plot without regard to economy of plant material; that is to say, he can use few replicates of large plots (see Section 51) and it is in such circumstances that calibration is commonly most effective. Of course, he will not be able to calibrate by means of previous performance because this will be unknown and the grower is unlikely to be willing to co-operate in taking records for some future trial, even if the delay could be tolerated. Nevertheless, he can make use of calibrating records like trunk girth, spread and shoot growth. Indeed, the differences between trees are likely to be so large that rough measurements of tree size, usually of little avail, prove quite effective in reducing variability.

Of course, such experimentation is by no means easy despite the help of calibration. There is usually a lack of trained helpers and their apparatus of weighing and recording implements, while distance can make economic use of labour impossible on account of travelling time. Above all, there is the difficulty of adequate supervision. The grower, however co-operative he may be—and some are very co-operative indeed—has almost certainly had no research experience and may do the most alarming things in the best of faith. Thus, while spraying his own trees he may spray the experimental ones as well and cut across the whole object of the experiment; nor is this danger confined to spraying but extends to all routine cultural operations. It follows that the experimenter and the grower must have a clear understanding and this they will never have if the grower is not visited often for a chat and a look round.

One special source of difficulty centres on picking. A grower wants the best price he can get for his fruit, so, when it is ripe or the market is especially favourable, he wants the fruit at once without waiting for careful picking of trees plot by plot followed by lengthy weighing. More trials on commercial holdings have failed on this account, probably, than on any other.

Allied to this difficulty is the question of compensation if some of the trees are rendered less fruitful than they would have been. Human nature being what it is, it would be surprising if grower and experimenter always saw eye to eye in this matter, but a lot of disputes would be avoided if designers would make a point of including a treatment to consist of those cultural operations the grower himself favours. It has several advantages. For one thing, it keeps the trial close to the local practices upon which it is hoped to improve; also, it provides a reliable test of the loss of crop occasioned by any treatment that raises a grievance in the owner of the trees and one that both sides can readily accept as a fair arbiter.

Finally, a grower who co-operates in an experiment must not only be visited often to see he does nothing wrong, but he must be treated as a collaborator and an ally. In particular, he will want to know what results are coming out of the trial and will appreciate a few figures even if they are hedged about with the reservations and caution commended in scientific circles. There is, after all, a certain prestige and reputation for enterprise associated with an experiment—but only if the grower has something to say when his friends ask about it.

It will be seen that this Section does not really fit into any chapter but the reason for it is worth noting. Experiments on outside plantations depend upon two pillars that are very diverse in nature; one is calibration and the other is tact.

45. The Method of Adjustment by Neighbouring Plots

In this method an expected performance of each plot is derived from that of its neighbours. Thus, if a plot is in the midst of others that are growing poorly its own poor growth is

unremarkable, but good growth would strongly suggest that the treatment given to it has been especially effective. The method thus has affinities to calibration in that the performance of each plot is judged in relation to an expected figure. On the other hand, like blocking it makes use of spatial patterns, not temporal ones.

The method here to be described is more general than the original form given by Papadakis (1937). Bartlett (1938) has established that the method will not give results seriously in error provided blocks are large, but its value as an experimental technique with perennial plants has always been problematical. The reason may well be that it has been used in wrong contexts. As is now recognized (Section 14), close associations in performance between adjacent plots is not to be expected when trees are young and it may be overborne by competition when they are old, but there must be many instances where trees have been in position long enough for their performance to be determined largely by local conditions and in which the spacing is such that competition is not serious. It is to be expected that the method would work well in such cases. Also a difficulty facing experimenters in many parts of the world is the patchiness of soil and the method should be at its best in such circumstances. Abeywardena (1964) used it with coconuts. Also, the results of Parker (1942) with citrus suggest that there can be considerable similarity between the performances of adjacent trees.

To illustrate the approach, the following figures are taken to represent yields from a group of plots in the corner of a trial, the letters representing treatments. The experimental design is for this purpose immaterial.

B, 60;	A, 40;	B, 56;	C, 45; . . . . .
C, 51;	C, 48;	A, 42;	B, 57; . . . . .
A, 47;	C, 50;	B, 57;	A, 39; . . . . .
B, 67;	A, 46;	C, 47;	B, 56; . . . . .
.	.	.	.
.	.	.	.
.	.	.	.

The next step is to work out treatment means over the trial as a whole. In the present example it will be assumed that these come to 45 for A, 65 for B and 50 for C, then it is evident that this corner of the field is a patch giving low yields, though to different degrees. The differences for each plot between actual figures and treatment means are:

B, −5;	A, −5;	B, −9;	C, −5; . . . . .
C, +1;	C, −2;	A, −3;	B, −8; . . . . .
A, +2;	C, 0;	B, −8;	A, −6; . . . . .
B, +2;	A, +1;	C, −3;	B, −9; . . . . .
.	.	.	.
.	.	.	.
.	.	.	.

It is now possible to get an idea for each plot how it might be expected to behave on the basis of the performance of its neighbours. The adjusting value for the second plot in the second row is $-7\ (=-5+1-3+0)$, the sum of the differences for its four neighbours, two along the row and two across it. Similarly the adjusting value for the third plot in the third row is $-12\ (=-3+0-6-3)$. The method breaks down for plots on the edge, but the method here is to take such neighbours as do exist and to adjust the numbers. Thus, the first plot in the second row should be given the adjusting value $-6\frac{2}{3}=\frac{4}{3}(-5-2+2)$, the corner plot $-8=2(+1-5)$ and the second plot in the first row $-21\frac{1}{3}=\frac{4}{3}(-5-2-9)$.

The adjusting value once obtained, it is convenient though not essential to add to each of them some constant number large enough to make them all positive. An analysis of covariance

(see Appendix IV) may then be worked out, the actual figure for each plot being adjusted by the values just derived. In carrying out an analysis of covariance it is usual to reduce the degrees of freedom for error and for total by one, but Bartlett (1938) has given reasons for thinking that the association between the dependent and independent variates in this type of analysis is such that the reduction is not enough, and he recommends that two degrees of freedom should be removed instead of one. The method of Wellman, Thurston and Whaley (1948) is similar to that of Papadakis but inferior, one difference being in the method of obtaining the adjusting values. That, in itself, is no criticism of their method. The precise formula whereby these values are derived from the differences from treatment means of nearby plots is not a point of principle—some prefer complicated expressions and others are content with simple ones, such as that suggested here. It is important only that the adjusting values should be derived in some objective manner independent of any preliminary opinions the experimenter may have formed of the probable effect of his treatments. The objection to their method lies in their avoidance of the analysis of covariance. The reason given in Section 41, namely, the need to keep clear of doubtful assumptions, applies with no less force here than in calibration; covariance is therefore essential.

CHAPTER 5

BLOCKS, PLOTS AND REPLICATIONS

" . . .—and a most curious country it was . . . the ground between was divided up into squares. . . ."

LEWIS CARROLL, *Through the Looking Glass*, Chapter II.

50. Size and Shape of Blocks

Since much of the variation in a plantation may be coming from sources that are not environmental, the study of size and shape of blocks is not as important as it is with annual crops, but even so it should not be neglected. The general rules are that blocks should be as small as possible and should be of some compact shape, i.e. square rather than long and narrow, but exceptions may be made in special cases. The real aim is that there must be as little variation as possible *within* them. Thus, on very hilly land "contour blocking" may be desirable, trees being assigned to blocks according to their altitude. That will be desirable because conditions at the top of the hill must be very different on account of depth of soil, exposure or drainage from those at the bottom, so greater uniformity within blocks will be obtained by dividing the land into long horizontal strips rather than neat squares, each of which may cover several contours. Thus Boubals (1957) recommended blocking grape vines by altitude and also by position on a terrace, whether inside, in the middle or outside. Similarly, if some part of the field has poor drainage or has different soil, a blocking system that takes account of the differences may well be preferable to one that does not, even though it uses blocks that appear to straggle when marked on a plan. To take an example, an apple rootstock trial was designed for a strip of trees, four rows wide, lying across the prevailing wind. In the event it was laid out in four randomized blocks, each occupying one row, so one block was on an outside row to windward, two were on inside rows and a fourth was on an outside row to leeward. Despite the length and narrowness of the resulting blocks it was considered that such a design, taking account as it did of the principal sources of variation, namely shelter and competition, was better than one in square blocks. Again, at a research institute that had to conduct experiments on irregular land, it was found that square blocks divided into long, narrow plots gave high variability and were expensive in surveyors' fees. Improvement in both respects followed when each block was formed along a drainage channel, all plots of the block being located similarly relative to the channel. The general rule for block shape is derived from the observation that neighbouring trees are usually more alike than trees far apart, but if there is reason to expect that resemblances will follow some other pattern, as in the cases just cited, the blocks should be formed accordingly. Some interesting examples are afforded by Agrawal (1961a).

The example of the apple rootstock trial also illustrates one useful property of blocks, namely, their ability on occasion to remove headland effects by arranging that all plots of a block shall be on the outside or, alternatively, all on the inside. Another example, this time from a cherry trial planned at East Malling but not planted, is illustrated in Figure XXII, where the effects of three unguarded headlands are eliminated by the blocks. A weakness here is the exposure of the two plots in the bottom corners to two headland effects, only one of which is eliminated, but that is probably not serious and could, if necessary, be rectified by the use of a pseudo-variate (see Section 81). As with the designs discussed in Sections 22, 24 and 32, if outside guards are to be omitted, it is essential that each headland should be homogeneous in its effect, i.e. there must be no buildings, trees, etc., that affect some of the outside trees on that headland but not others. Roosje (1963) similarly has pointed out the need to take account of windbreaks, which can often be done by blocking. Simpson (1936) likewise has drawn attention to the variation brought about by rows of pollinators.

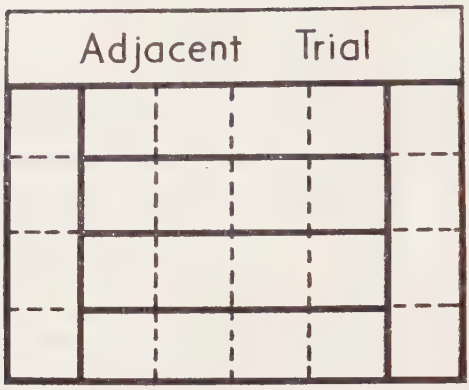

FIGURE XXII. A system of blocks to eliminate headland effects.

It is also desirable that all blocks of a trial should be similar in shape and area, though that is not always possible. Usually they will be of the same area anyway, because each will contain the same number of plots, but exceptions do occur. For example, if the land has obstructions, such as outcrops of rocks or a winding stream, spaces may have to be left that will increase the surface of some blocks but not others. As to shape, there is a temptation sometimes to form most of the blocks in some compact manner and to assign the residue of plots to the last block, no matter how diverse they may be. An example is illustrated on the left-hand side of Figure XXIII, the right-hand side suggesting a better method.

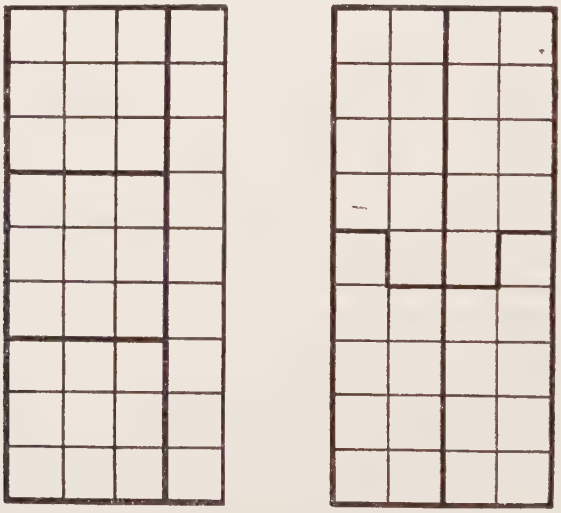

FIGURE XXIII. Alternative methods of forming blocks.

Numerous experimenters have had the idea of forming blocks on the basis of the past performance of the trees (e.g. Koch and Rigney, 1957), but there are strong reasons for preferring calibration. For one thing, blocks so formed may indeed allow for past sources of variation, but covariance is likely to be better, because it will take account of *all* variation in the independent variate, whereas only *some* variation can be associated with blocks, the

rest being within them. For another, blocks are needed on account of possible future variation as well as what is past and a dry year or a frosty one may impose patterns quite different from those experienced previously. Such patterns are more likely to be associated with compact blocks than with straggling ones. Also, if the calibrating characteristic proves to be irrelevant, a covariance adjustment represents little loss but a distortion of the design could be serious. Finally, only blocks of contiguous plots make convenient administrative units. For these reasons it is inadvisable to remove known environmental variation by blocks instead of covariance. Sen (1963a, b) points out that calibration is most effective when the blocking system is poor. He found with crops of tea that there was little to choose between calibration and blocking as a means of allowing for known past differences.

In order to conserve degrees of freedom it is possible where plots lie side by side to use covariance on linear and parabola trends, etc., either in addition to blocks or instead of them (Stevens and Fraga, 1951; Federer and Schlottfeldt, 1954; Outhwaite and Rutherford, 1955). Before adopting such a plan the experimenter should be very sure that the trends he proposes to exploit really do exist.

A blocking system having been adopted, good use should be made of it throughout the whole period of the trial, as described in Section 20. In order to make blocks of such a size that they correspond to a day's work in tapping rubber or plucking tea or other operations, confounding or the adoption of non-orthogonal designs may be necessary.

There are then two purposes of a blocking system. One is to reduce the experiment to convenient administrative units; the other is to control local variation and there can be conflict between them. However, where variation is largely random the blocks can do little to reduce variability, so administrative considerations can be given full weight. On the other hand, in a smaller experiment there may be no administrative reasons for dividing it and blocks, if needed at all, can be chosen to allow only for local control so far as that is necessary. There can be no general rule in these matters but only a recognition of all relevant considerations. Above all there should be no misconception that the formation of blocks is an exercise in geometry. For administrative purposes it is indeed often helpful to have compact, rectangular blocks, but they can be of any shape that is expected to be effective.

51. Size and Shape of Plots

Just as there are no general rules for blocks so there are none for plots and for very similar reasons. If random variation predominates, the form of plots matters as little as the form of blocks, but if there is local variation to be controlled everything goes by opposites, because blocks are required to pick up differences whereas plots within a block are required to be as uniform as possible. Hence, if blocks have to be compact to allow for unpredictable irregularities, plots should be long and narrow so that each covers as wide a range of conditions as possible. Again, if there is an environmental trend, as on a slope, plots should be at right angles to the blocks.

As to size of plots the leading reference is that of Fairfield Smith (1938) though there have been a number of developments as set out in Section 14. However, in so far as they relate to efforts to combine Smith's law for local variation with that for random variation they have no bearing here, because the addition of sources of random error dilutes the local variation without altering its nature. More serious are the doubts that have been raised about Smith's law itself, e.g. those of Rives (1969), who found that larger plots of grape vines may be more variable than smaller ones, and those of Matérn (1970), who pointed out that the law implies some unlikely relationships between the performance of plants and the distance between them. There is now a substantial body of evidence that economy of land is best effected by using small plots, but that more information per replicate is obtained from large ones. Thus, if an apple trial is being designed without guard rows and it is intended that the significant difference for crop between treatments shall be about 20 per cent of the general mean, the experimenter

might be faced with the choice of using eight replicates of one-tree plots, four replicates of three-tree plots or three replicates of five-tree plots (Pearce, 1949). His decision would depend entirely upon circumstances. If he were planning the trial for a research station with the usual shortage of land but a sufficiency of trained staff, he would choose the smallest size of plot; if he were planning for a grower's plantation with many available trees but no skilled experimental assistance, he would choose the largest. There is no general rule in the matter. Reports of uniformity trials that purport to inform their readers of the "best plot size" for a certain species are merely misleading.

In deciding upon a suitable plot size, the principal consideration is the one just given, i.e. to economize land, use small plots; to economize skilled labour, use large ones. There are, however, a number of subsidiary points to consider. Thus, plots must be sufficiently numerous to give enough error degrees of freedom—say 15 and, if possible, 25. Also, plots must not be so small as to give a standard error greater than 30 per cent of the general mean, though that can be considered only in the light of past experience. Also, a transformation (Section 73) will often help matters. Further, plots should be of such a size that their crops can be conveniently weighed in the field, difficulties being caused by quantities so large as to need numerous weighings and by quantities so small that the diminutive scales required are unusable in anything except a complete calm.

An important consideration determining plot size must always be the kind of record that it is proposed to make. For example, in a nursery trial it is often desirable to know the proportion of plants that will root or blossom in the first year or even the proportion that can survive. In such cases it would be ludicrous to use plots consisting of only a single plant, for then the data can only read 0 or 100 per cent, nor would the situation be much helped by using two plants for that would only increase the possibilities to three, namely 0, 50 or 100 per cent. If discontinuity of variate is to be avoided a plot must consist of at least ten plants and even that is minimal. On the other hand, a determination of shoot growth on plants that do root successfully can be made adequately on a small number. If shoot growth were all that mattered, much smaller plots could have been adopted in the first place. Again, when it is proposed to sample leaves from each plot of a field trial throughout the season, the plot size has to be large enough for the loss of leaves to be unimportant. Otherwise the data may show a trend which has nothing to do with the treatments but results from the progressive defoliation. On that account multi-tree plots may be adopted, whereas single trees might have sufficed if only one set of leaf samples had been required and that towards the end of the growing period.

Plot size must also be considered in the light of possible losses. Here it should be explained that, in the analysis of data, completely missing plots are manageable provided there are not too many of them; at the other extreme, no irreparable harm is done if one or two plants die in a plot of, say, twelve or more. Difficulty is, however, caused by a plant being lost from a medium-sized plot. Consequently, if losses are expected—and with some species they are inevitable—there is a case for using either large plots that can stand a few losses or small ones that will be completely wiped out, but for avoiding medium-sized ones in which losses can be very awkward. Admittedly, however, there are exceptions. Thus, at East Malling it was established practice with plums and cherries to use two-tree plots, but those species are distinctive in that most losses occur early. Consequently, it is possible after quite a short period to decide upon the course of action: either (1) to regard all plots from which one or two trees have been lost as completely missing, (2) to reduce all plots with survivors to one tree by the random elimination of a tree where both have survived and so to accept a few completely missing plots, or (3) to replant all losses simultaneously and use a pseudo-variate (see Section 81).

Two important general points remain to be noted. It sometimes happens that (1) the treatments are being applied directly to the trees as individuals, no guard rows being needed, and (2) each tree must be recorded separately. In such cases the skilled labour needed is

directly proportionate to the number of experimental trees which, in the absence of guards, is directly proportionate to the area of the experiment. That is to say, there is no longer any conflict between the economizing of land and of labour and there is a strong case for using the smallest possible plot—one-tree plots, in fact, unless there is a good reason to the contrary.

A final general point concerns the effect of guard rows. The smaller the plot the better is the use of land but the greater the wastage due to guard trees; it may not be at all clear which of those two effects will be dominant. That this difficulty has received such little attention is rather surprising in view of the extensive use made of guarded rows, but it is clear that small guarded plots should be used with reserve. Incidentally, square plots need fewer guards than long ones of the same area.

All this will help when experience has once been gained with a crop. It will suggest when to make plots smaller and when to make them larger, but it will not help if the investigator is working with the species for the first time. Here the recommendations of Section 52 may be of help, since they are given by people who know something of the various crops. Even though some of the trials are open to criticism and though the conclusions reached need not be adhered to as rigidly as the writers sometimes appear to suggest, they do represent a reasonable sort of size for the beginner to adopt. Actually, plot size may be varied to a surprising degree if replication is adjusted with it. The largest plot the writer can recall was used by an irrigation engineer, who wanted to know what advantage would be gained by setting up small dams to prevent the water of mountain streams running down the hillside into the river at the bottom. The experimenter proposed choosing a number of valleys and in each selecting one side at random for his dams. Since the government of his country had already voted a sum of money for the erection of such works over a stated area and since they had intimated that they would require evidence of agricultural improvement before voting any more, the engineer's proposal appeared completely sound even though he was suggesting plots of at least 100 square kilometres area. (He also proposed a sampling scheme for the produce of farms before and after the works had been completed to provide data.) At the other extreme, plots consisting of one leaf or one bud are common and there is no objection to plots of a single cell if someone can devise a technique for using them. Size as such is of no importance so long as the plots are acceptable on practical grounds, are independent of one another and reasonably uniform. Largely on practical grounds Narayanan (1965) recommended using plots of medium size and there is a lot to be said for such a policy.

One of the most illuminating studies is that of Laycock (1955), whose paper presents not only yield data for individual units but contour lines. He was working with tea on irregular land, using square blocks, and there were no differential treatments, so plots could be formed at will. First, he used long, narrow ones, the directions being the same in all blocks, and results were satisfactory. Then he used a second set, similar to the first but at right angles, and again results were satisfactory. Then in each block he chose the set of plots that lay down the slope and across the contours; results notably improved. An important conclusion emerges. With all plots going the same way, some blocks must have been making large contributions to error and some small, thus invalidating an assumption of the analysis of variance that all plots are equally subject to error. The consequences will be examined further in Section 73; the point here is the need for careful consideration of the plot structure. A lot depends upon circumstances. Thus, Páez (1964), working with coffee, found orientation to be important, but some others have not found it to be so. Paardekooper (1966), whose work on variability in rubber well repays study, noted that the best plots appeared to be those that went right across the experimental area, a result possibly of the ease with which such plots can be worked. A good summary of the considerations in choosing a plot size with rubber has been given by Westgarth (1953).

It is not essential for plots to be rectangular, though it is often convenient. Some have advocated that they form segments of a spiral (Horsfall and Rich, 1949), while others, especially those who have to work on peasant farms, have found it easier to use circular ones (Cunningham

and Burridge, 1959). As Hänninen (1951) points out, all the apparatus needed is a peg and a length of cord.

Some diseases show a tendency to spread along rows rather than across them. Where that is expected to occur it must have an effect upon the orientation of plots. Despite certain general recommendations that in such cases plots should be as square as possible (Edgar, 1938) or should be disposed across the rows (Baker and Baker, 1950), the present writer believes that the decision must depend upon the purpose of the trial. If the intention is to test some cultural treatment, the expected disease being no more than a nuisance, it is obviously wise to dispose plots across the rows so that they are all affected to about the same extent if a row does become diseased. If, on the other hand, the disease is itself the object of study, it would appear better to arrange the plots along the rows in order to confine an outbreak to the plot in which it started.

Despite the economy often afforded by small plots, reduction should not go too far. Rich (1952) pointed out the advantages of using half-leaves instead of whole ones when assessing fungicides, his reason being the similarity between the two sides of the lamina. If the block is to be the whole leaf, the advantages are obvious provided that treatments are few. (If a large number of treatments require comparison, studying them in pairs will reduce efficiency so much that the smaller variance will have been purchased at too high a price.) If, on the other hand, the block is made larger than one leaf, the differences between leaves will still swell the error, but some comparisons will be made better and others worse than the average according to which pairs of treatments have been assigned by randomization to the same leaf.

52. Review of Uniformity Trials, etc.

Once upon a time every good experimenter conducted a "uniformity trial". He took an area to which no differential treatments had been applied, split it up into small units, each of which was recorded separately, and then he worked out the error variances he would have encountered had he used plots of various shapes and sizes within different kinds of block. By such methods much valuable information was obtained, upon which many notable research centres based their experimental techniques, usually with some success. Uniformity trials are no longer fashionable and it will be helpful to see why, but no one should decry the efforts of the early investigators.

Nevertheless, few people conduct such trials nowadays and one reason is justified scepticism that there is a "best plot size" to be discovered. Too much depends upon the balance of environmental and other variation, which itself is not constant but is different at different ages of tree, at different sites and with different kinds of plant material. Further, it was once supposed that as plot size increased the error variance fell until it reached a limit. In formulating his law, Fairfield Smith (1938) showed that to be little more than an optical illusion, the apparent shape of the graph being quite different if scales of measurement are altered. Once it was clear that variance mostly continued to fall, the problem became one of finding the optimal combination of replication and plot size, which is primarily a question of convenience and cost.

There are also certain technical objections to the old-time uniformity trial. For one thing, if a good range of plot sizes is to be studied and if even the largest are to give enough degrees of freedom, there should really be about 300 units in all and not all experimenters managed to attain so many. For another, by a curious lapse of perception, many of the investigators kept their block sizes constant while altering the size of plots. In fact, if plots are changed in size, blocks should be also.

Nevertheless, a list of published uniformity trials may be useful. The following is in alphabetical order of species. With one exception the data were measures of crop. Those references marked with an asterisk present the original data.

Apples	* Batchelor and Reed (1918); * Strickland (1935); Pearce and Thom 1951a); Schultz and Schneider (1955); Neumann (1958b); Abraham, Hoobakht and Nejad (1971)
Avocados	Jones, Embleton and Cree (1957)
Black currants	* Hatton, Grubb and Knight (1925)
Cocoa	Cheesman and Pound (1932); Jolly (1942); Pearce and Thom (1951b); Van der Knaap (1955); Páez and Siller (1963); Esquivel and Soria (1967); Lotodé (1971)
Coconuts	* Joachim (1935); Pieris and Salgado (1937)
Coffee	Gilbert (1938a, b); Van der Knaap (1955); Páez (1964); Awartramani (1965)
Grapes	Gladwin (1924); Strickland, Forster and Vasey (1932); Covas and Christensen (1945); Christensen (1948); * Geisler and Staab (1958); Grácio (1969); Hegediüs (1969)
Hops	Keller (1949)
Lemons	* Batchelor and Reed (1918); Jones, Embleton and Cree (1957)
Oil palms	Webster (1939); Ollagnier (1951)
Oranges	* Batchelor and Reed (1918); Jones, Embleton and Cree (1957)
Papaws	Younge and Plunckett (1969)
Peaches	* Strickland (1935); McHatton (1947a, b); * Shah (1950); Schultz and Schneider (1955)
Pears	* Martinez-Zaporta (1952) (worked with size records of the trees, not crops)
Pecans	Sharpe and Blackman (1950)
Pepper	Anon. (1958); Abraham, Khosla and Agarwal (1969)
Pineapples	* Magistad and Farden (1934)
Raspberries	Hoffman (1930)
Rubber	Paardekooper (1966); Iyer (1968)
Strawberries	Wilcox (1927); Edgar (1938); Taylor (1951b); * Baker and Baker (1953)
Tea	* Eden (1931); * Laycock (1955); Sen (1963b)
Walnuts	* Batchelor and Reed (1918)

The difficulty with all this work lies in the generalization of conclusions. An investigator wants to know about the variability of the species he works on and accordingly he searches the literature for what information is available. In fact, the conclusions may not depend much on species. For example, someone may have reported that long, narrow plots were better than square ones of the same size, but that must surely indicate a fertility gradient rather than a specific characteristic. Also, where trees are growing well they may be less susceptible to small environmental differences than others that are struggling to survive, so results obtained with a species in one country may not be repeated in another or on different soil. Not least, as trees become older their performance will increasingly reflect continuing environmental effects, but the stage at which such effects dominate will depend both upon their magnitude and the state of development of the trees. If plants have grown slowly and with difficulty, they will take longer to settle into established patterns. In fact, it is better to consider evidence from all species together and to discern general rules for plot size and shape, which should then be applied with discretion. There is no reason to suppose, for example, that there is one set of rules for apples and another quite different set for oranges. On the other hand, the old saying that "circumstances alter cases" applies all the time.

Even more difficult are generalizations about the level of variability. Here an experimenter must know what happens in his own conditions. Very little is to be gained by studying the literature.

It may nonetheless be useful to look at the more extended studies and see what came out of them. Martinez-Zaporta (1952) worked with pears grafted on quince and recorded trunk circumference and height. He found remarkably little effect of plot size or shape, but he did find different variability on different rootstocks. Neumann (1958) studied crop, trunk circumference and crown measurements of apple trees. Some of his values for Fairfield Smith's b exceed one, i.e. they indicate a fall in variability with increasing plot size that is more rapid than would be expected with random variation, which implies a negative correlation between adjacent trees, the result perhaps of competition. However, the greater number lay between 0·5 and 1·0. As Neumann remarks, with such high values there is little need to minimize block sizes by devices such as confounding, though he also points out that with his trials in general larger plots did not reduce in variability as fast as would be expected of a random distribution, i.e. b was in general less than one, its mean value being about 0·74. The whole study is most thorough and well repays attention. For all cases he favoured the smallest plot size practicable, but considered that to be largely determined by the object of the investigation and the experimental techniques to be used. Roosje (1962) favoured one-tree plots for stone and pome fruits, but noted that there are occasions when they are not practicable.

Chapas (1961) based his results on two uniformity trials of oil palm, using data for twelve years from one and for five years from the other, the quantities studied being number and total weight of bunches. He did not find plot shape to be of much importance, but the coefficients of variation did depend upon age, falling as trees got older and crops increased. Also with increasing age, for one trial, values of Fairfield Smith's b rose, indicating less association between neighbouring trees, not, as might have been expected, more as environmental effects built up. When crops were grouped into two-year periods the coefficients of variation fell markedly, as did values of b, indicating that the individual variation of trees was now relatively less.

Working with tea Dutta and Heath (1960) found different variances with different clones. Long, narrow plots proved to be better than square ones. Plots of 30–45 bushes were recommended because smaller ones would cause difficulties of weighing, but for some treatments larger ones might be required.

Of especial value are surveys of variability found with a given species in a range of conditions using plots and blocks of different kinds. For strawberries there is the work of Taylor (1951b), Ferguson (1963) and Holland (1971), while for black currants there is that of Garretsen and Kronenberg (1957), Ferguson (1963) and Holland (1969). However, even with such surveys the experimenter must consider his own conditions, which could well be different.

In thinking of variation it should be recalled that variances depend not only on the circumstances but upon what is measured. Thus, a trial able to show a 20 per cent difference in crop might be able to show a 15 per cent difference in trunk circumference but only a 40 per cent reduction in the incidence of some disease. Also, crops are not equally variable at all ages, nor are trunk circumferences, while the duration of the experimental period also is important.

These difficulties are presented, not to argue that no prior forecast of variability is possible, but to warn the reader against facile generalizations. In course of time each research centre builds up a body of local lore that will provide guidance and that is all that is needed, but care must be taken to use it intelligently. When an experimenter comes fresh to a project he may well find useful guidance in the results of uniformity trials and they can be supplemented by noting the errors reported in the papers of other workers. Admittedly, many experimenters do not have extensive libraries to hand, but they must all be in a position to read the more important papers relating to the species they study. If they would note details of experimental design and the precision of results as well as the results themselves, they would be less worried when it came to conducting experiments themselves. Also, if they would mention their own standard errors when they come to publishing, they would help those who come after.

53. The Use of Guard Rows

Guard rows serve a variety of purposes and their use should be carefully considered.

First they serve to eliminate the difference between inside and outside trees, which can be very important if the treatments are of a kind that can be expected to interact with a reduction in competition, as is the case with most that are applied to the soil, such as fertilizers or herbicides. Where no interaction is expected, a clever designer can often make use of blocking systems that consume headland effects, row-and-column designs being especially useful for this purpose, though, where split-plots are needed, it is often difficult to arrange for the effect of outside trees to be spread over both main plots and sub-plots. Where blocks cannot deal with headland effects, guard trees must be used.

Such guards may be described as "external" and more than one row of them may be needed because microclimate can change for some way into a plántation. In particular, windbreaks can be a nuisance. For one thing they can set up root competition that does not necessarily finish with the outside row; for another they can cause turbulence to leeward that is often worst some way into the plantation, usually at about five times the height of the windbreak and down-wind from it. This turbulence is less if the windbreak is not a solid barrier but rather resembles a slatted frame, because it then slows the wind rather than lifts it. Where the wind is lifted, there can be effects to windward as well as to leeward. Because trees grow in height and become thicker, a windbreak that starts with slowing the wind can later lift it and cause turbulence at an ever-increasing distance to leeward. Hence, conditions need to be kept under periodic review.

Another awkward hazard near an experiment is water. Especially in a dry climate a pond can have wide-reaching effects to which guard rows hardly provide an answer. There have to be so many of them that effectively the experimental area is pushed back to a more uniform environment. Care should also be taken with buildings. Anyone who has watched a field of wheat to the lee of a house while the wind is blowing will have noticed the erratic patterns of turbulence thus revealed. Needless to say, it is useless to try to allow for the outside effects by blocking if headlands are heterogeneous (see Figure XXIV).

It is the practice at some places to use special guard rows between blocks. There is, however, no need to guard the boundary between blocks more thoroughly than that between two plots in the same block, so the practice is quite mistaken.

Another need for guard rows arises when treatments are not applied directly to the plants themselves but, like applications of sprays and fertilizers, are intended to change the environment of the plants. There is then the danger of treatments affecting trees to which they are not applied, e.g. on account of spray drift or poaching by roots. Such guards are called "internal".

Such guards present no problems as long as they do not have to be doubled. The question then arises: What has to be done with the alley between them? If it is not fertilized, there could be two rows of stunted trees that provide no effective barrier against poaching roots, whereas if the fertilizer is extended from the experimental plots to the middle line of the alley much of the point of having double guards is lost. Similarly in a spraying trial, the choice could lie between leaving possible foci of infection or of risking spray drift on to experimental plants. With three guard rows the middle one can be treated differently on the two sides and there is still a full row on either side of it to give protection to the plots.

Similar to internal guards and sometimes scarcely to be distinguished from them are "buffers", i.e. rows of trees left to prevent differently treated trees from affecting their neighbours. For example, in a trial of insecticides a reduction of the standard spray schedule may lead to the build-up of a mobile species that could endanger the whole experiment. Quite apart from internal guards to make sure that the experimental trees receive the sprays intended for them and no others, buffers may be needed in the plots themselves, only an area in the middle being recorded for fear that the parts of the plot nearer the edges may have been invaded by insects from the plots round about. Sometimes the situation is described

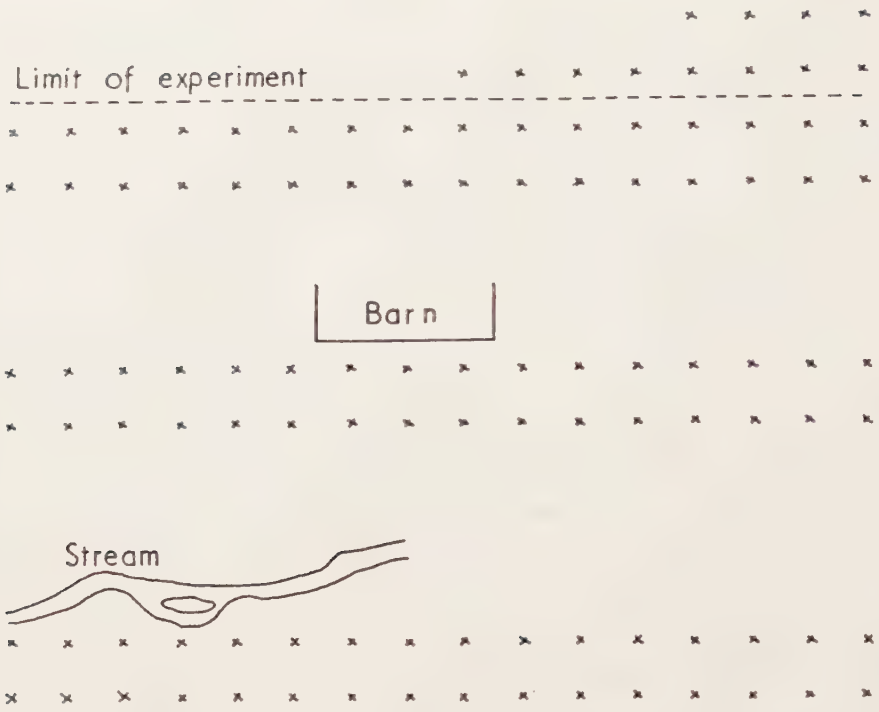

FIGURE XXIV. Examples of heterogeneous headlands.

by saying that the recording plot is smaller than the spraying plot; the difference is only one of nomenclature.

Buffers are sometimes needed even when there are no internal guards. That can be the case in a rootstock trial. There can be no fear that the treatments, i.e. the rootstocks themselves, will affect trees not grafted on them, but it could well be that a vigorous tree will crowd its less vigorous neighbour. In that case it is a little difficult to suggest what a single buffer row should consist of. If it is the same as either of its neighbours its job is only half done, while to use a standard rootstock of medium vigour for all guards satisfies no one, especially if the chosen rootstock is likely to be an ineffective competitor to the more vigorous experimental trees while overpowering the less vigorous. Unless double buffer rows are used there seems to be no solution and they are very expensive. For that reason extremes of size are best avoided, not only in the trial of rootstocks but of anything else.

Alternatives to guard and buffer rows exist and can usefully be considered. For example, an alley may be left, either by omitting a row of trees or by increasing the planting distance. As a barrier against the spread of pests and diseases the device is often valuable, while it also facilitates the erection of screens when spraying. In fertilizer trials it has the disadvantage of altering the area of land available to each tree, but here there is the possibility of placing barriers in the soil (Goode and Marchant, 1962; Marchant and Boa, 1962). Another device sometimes adopted is a deep guard furrow, frequently renewed to cut roots that try to pass it.

In Figure XXV are shown some different kinds of guards, differing mostly in degree.

Certain observations on guard trees merit special mention. Thus, Vernon (1968) found substantial edge effects in a fertilizer trial on cocoa. Haines, Crowther and Thornton (1954) revisited a former rubber experiment after its abandonment for some years and found marked

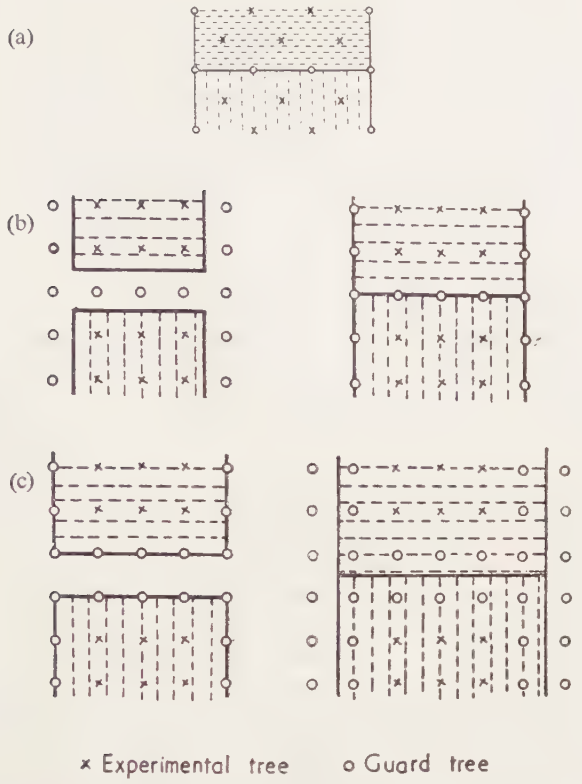

FIGURE XXV. Different types of internal guard rows. (a) Half-guarding. (b) Single-guarding with and without alley. (c) Double-guarding with and without alley.

poaching of fertilizers over guard rows. Monzón, Haddad and Figueroa (1969) pointed out that where spacing is wide the external guards are not very different from the rest. For passion fruit they found that to be so when rows were 3 metres or more apart. Proebsting (1942) noted that outside trees of apricot, peach and pear were sometimes more fruitful than inside ones and sometimes less so, but were only rarely comparable.

54. Degree of Replication

Replication serves two purposes: (1) it makes possible an estimate of the residual variability of the experiment by providing error degrees of freedom, and (2) it enhances the estimates of treatment effects, which otherwise would be based on single plots.

Although increased replication does lead to better determination of treatment means, it is not wise to base precision on that alone. For one thing, it is often disappointing in its result, e.g. replication has to be increased fourfold to halve the standard error. For another, standardization of experimental material and calibration are usually more effective. In deciding the degree of replication called for, the possibility of losing plots must be borne in mind.

However, no decision can be taken on the degree of replication without reference to the purpose of the experiment. Perhaps it is intended only to show large differences. If the cost of applying a treatment is high, small differences are of no interest because the gain will not pay for the increased costs. Perhaps the experiment could usefully show small differences as in the case of introducing a new rootstock or virus-tested material. Here, once stocks have

been built up, the continuing cost of using one strain rather than another is small and a quite unremarkable increase in cropping would pay for itself. Further an endemic virus disease that reduces yield by 5 per cent can do more harm to a farmer's bank balance or a nation's economy than a condition with large effects that occurs rarely and can be readily exterminated when it does. Nevertheless, there is an upper limit to the size of experiments here too and it arises from alternative uses for the research worker's time and the resources available to him. It might indeed be very worth while to raise productivity generally by 5 per cent by deciding which strain is to be preferred, but an experiment intended to establish so small a difference might necessarily be so large as virtually to preclude research on any other topic.

A warning is needed about preliminary trials. Experimenters sometimes excuse a scrappy little investigation on the grounds that it is "only preliminary". If they mean that it is concerned only to establish feasibility, they may well be justified. Some treatments cause extensive damage; some graft unions rarely "take" and so on. If there is any doubt on these matters, it is wise to check before embarking on a large-scale trial. However, many preliminary experiments need to be very elaborate indeed. After all, if it is the intention that no further work shall be done on a certain line if it does not appear promising but efforts shall be concentrated on something else that shows a larger difference in the preliminary investigation, the decision is indeed a momentous one and it should not be left to the vagaries of random effects. Before abandoning a line of possible advance, there should be some confidence in the power of the test used, i.e. its ability to detect an effect if there is one (see Section 12). Such confidence cannot be reposed upon a poor experiment.

A special problem is presented by factorial designs, which sometimes provide "hidden replication", as it has been called. Thus, if a trial has six replicates of eight treatments, made up of one factor at four levels and another at two, and if there is no interaction, the first factor has in effect 12 replicates and the second has 24. It would be wrong to rely upon such replication because the inclusion of the two factors in one trial implies a belief that they may interact, but in multifactorial designs with high-order interactions, which are not themselves under study but merely arise from the nature of the design, it is often quite reasonable to assume that there will be no last-order interaction and to reduce actual replication accordingly, relying upon the hidden replication to make good the deficiency.

Such hidden replication has a bearing on the design of preliminary experiments. One good way of finding which of a range of possible lines should be adopted is to lay down a 2^n factorial experiment with each factor representing two extremes of whatever treatment is being considered, e.g. no potash and high application, no cutting back versus severe and so on. Actual replication need not be high. A single replicate may suffice or even half a replicate (Finney, 1945b, 1946d, 1950). Suppose, for example, that seven lines of advance have been proposed then a 2^7 design of a single replicate of 128 plots could be excellent. There would be 7 degrees of freedom for main effects, 21 for the two-factor interactions and 99 for error, i.e. all the other interactions. The experiment would show if factors had to be considered in groups, i.e. an interaction between A, B and C in pairs would give a fairly strong hint, while 32 plots for each cell of a 2×2 table of means is, if anything, excessive. In fact, many would be satisfied with a half-replicate of 64 plots. Also, most people would want blocks to be introduced, so possible complications are endless. The point being made here is, however, simply to indicate the potentialities of a multifactorial design that makes good use of its hidden replication. Preliminary trials should not be rough and inadequate; on the contrary, the simplicity of the questions they are intended to answer allows for great power and directness in the design.

55. Formulae for Determining the Number of Replicates

In the literature two methods have been given for determining the replication needed to show a significant difference of a certain amount. In the one, which has many forms, the

argument goes like this: the error variance may be expected to equal s^2 (a figure based on past experience as described in the last Section). With an effective replication of R, the smallest difference between treatment means that will appear significant is $D = ts\sqrt{(2/R)}$, where t is derived from tables and depends upon both the level of significance required and the number of degrees of freedom in the error. Equating D to the figure desired will give a value for R.

The method is not wholly satisfactory. It will be seen that the above value of D is an average figure, depending as it does on an average figure of s. In fact, there is a half-chance that in any particular instance s will be above expectation, just as there is a half-chance that it will be below. Consequently, if a value of D is selected as representing the sort of sensitivity desired and a value of R adopted accordingly, there is a half-chance that variation in s will prevent the selected difference, if it occurs, from being shown significant. Nevertheless, in making a general recommendation, the first method is correctly used as giving an average figure for sensitivity over future trials if the value of R is adopted.

In the second method (Cochran and Cox, 1950) a difference, D, is taken and associated with a probability P'. The investigator says in effect: If a difference, D, should arise, I wish it to be detected on a proportion, P', of occasions, P' being greater than or equal to 0·5. The equation now becomes $D = s(t+t')\sqrt{(2/R)}$, the symbols having the same meaning as before. t', which is new, is obtained from tables in the same way as t, the degrees of freedom being those for error and the significance level being $2(1-P')$. It will be seen that if P' is made equal to 0·5, the second method reduces to the first, t' equalling zero.

Thus, suppose s^2 is expected to equal 400 and it is proposed to compare six treatments in randomized blocks. It is desired that if a difference of 25 should arise between treatment means there should be a probability of 0·8 of detecting it at the $P = 0{\cdot}05$ level. Suppose nine replicates were used. That would give 40 degrees of freedom to the error, making t equal to 2·021 (for probability, P) and t' equal to 0·851 [for probability, $2(1-P')$], so the value of D would be $20\,(2{\cdot}021+0{\cdot}851)\sqrt{\frac{2}{9}} = 27{\cdot}1$, and evidently that is not replication enough. Further investigations show that eleven replicates would be needed to make D less than 25, a result that could have been obtained from the table of Cochran and Cox* (1950). Reference should be made to the same place for the replication needed in one-sided tests, i.e. where it is assumed that a certain treatment cannot be worse than the control, though it might be better (see Section 71).

In deciding upon the number of replications required, it is permissible to take into account any hidden replication arising from high-order interactions assumed to be non-significant. Also, the possibilities of losses should always be borne in mind. Agrawal (1961b) has noted the incidence of death among experimental trees of various species in the conditions of north India. Other research institutes could no doubt search their records and compile similar figures for their own conditions.

* These writers use σ for s, P for P' and δ for D, a notation avoided here, partly from a desire to keep Greek letters for parameters and partly to avoid confusion with P used to indicate a significance level.

CHAPTER 6

TRIAL DESIGN IN RELATION TO THE SUBJECT OF INVESTIGATION

"Every purpose is established by counsel."

Proverbs, xx, 18.

60. Fertilizer Trials

There are two broad classes of fertilizer trial which may be termed "definitive" and "tentative". In definitive trials certain treatments are applied and continue to be applied even though the trees show signs of distress, because the purpose is to elucidate the precise effect of certain elements, or the lack of them, or the effect of applying fertilizer in a certain way. If the effect turns out to be stunted growth or abnormal leaf development the fact needs to be established and recognized for the future. In tentative trials on the other hand the object is to find the best fertilizer programme for growing that sort of tree in those conditions. Consequently if a treatment gives rise to deleterious effects it is discontinued and possibly another is substituted for it. The two kinds of trial are so different that an experimenter must be clear from the start which kind he is conducting. As long as all treatments appear to be advantageous, though perhaps in varying degrees, there is no conflict, but if problems arise it is important that decisions shall be both reasonable and consistent one with another.

Another problem concerns initial levels. In a field that has been well fertilized applications of an element, even an important one, may have little effect, whereas in other conditions the improvement could be most marked. For that reason, quantitative work on optimal levels of fertilization is difficult unless conditions can be defined, which is often a difficulty in itself since chemical determinations of soil and leaf tissue rarely tell the whole story. Such quantitative work could conceivably take several forms. For example, as Finney (1953b) has pointed out, there comes a time when the increased cost of fertilizer is no longer repaid by the increase in yield, a line of thought that calls for the fitting of curves relating yield to level of application (Mitscherlich, 1934; Pimentel-Gomes, 1951a, b, 1953). Again, in many industrial processes it is usual to fit response surfaces in which quantities like temperature and pressure are assigned various combinations of values to estimate the combination that gives the maximum response in terms of a desired product. There is a clear application to changing levels of N, P, K, etc. to obtain the combination that gives maximum crop and the statistical techniques are not difficult (see, for example, Cochran and Cox, 1950, not first edition), but again there is the problem of initial levels. Nevertheless, there remain parts of the world's surface where conditions are fairly uniform and in pot experiments they can be made so. Nelder (1966) used such methods with annuals in the Fen country of eastern England and they could well be used with perennials in suitable circumstances.

Initial levels are also, of course, involved in the generalization of conclusions because a result found at one place does not necessarily apply at another. In the early literature there is the amusing story of a man who began by reporting that phosphate increased the yield of potatoes and ended, after repeated criticism, by asserting only that it did so in Old Warren Field at Rabbitsbury, thus leaving all generalization to the reader. It is significant that the story was told of a fertilizer trial. If the experimenter had found out something about the differentiation of fruit buds or the development of shoots, it would have been assumed that the phenomenon lay so deep in the functioning of the plant that environmental differences could change it very little, but with fertilizer there is no such assumption. It is also significant that dispersed methods of experimentation (Panse and Abraham, 1959) have been developed with fertilizer trials chiefly in mind. If a fertilizer recommendation is to be made, it needs testing on a range of sites, either to measure the mean benefit or to discover the kind of locality that it fits. Alternatively, or in addition, a soil survey may help, though it must be coupled

with some idea of what happened in the soil of the original experiment. Also, it is helpful if diagnosis and remedy can be seen as separate problems. It is sometimes possible to show that a plant needs an element; in the old days by injection (Roach, 1938; Roach and Roberts, 1945), though that appears to be out of favour nowadays, or by a foliar spray. Once it is clear what element may relieve the chlorosis or the stunted growth or other trouble, a field trial may be needed to discover how best to make it available to the plant, especially if the element is one that tends to get locked up in the soil.

It should also be borne in mind that the nutritional needs of a perennial plant may well vary with time. Accordingly it may be necessary to have two or more factors for the same element to represent applications at different periods, e.g. during establishment, from first major crop, etc. Ideally their interactions should not be confounded because they could be important, i.e. early application of a fertilizer might obviate the need for large dressings later or it might lead to well-grown trees that require the larger amount.

Experiments involving fertilizer mixtures usually run into the problem of what exactly is being studied. Is a 30 : 30 : 40 mixture of N, P, K really to be preferred, or would it be better to use 25 : 40 : 35? The difficulty is to know the basis of comparison. If cost, the relative prices of the two mixtures will not remain constant; if weight, a kilogram of the one is not really comparable with a kilogram of the other, especially if inert filling is added to aid distribution. (The position is different if the question concerns the form in which a certain element is to be applied because there is then a rational basis for declaring two treatments equivalent.)

In a tentative design care must be taken to ensure that new factors can be added as new questions arise. In fact, such a trial may well be started with no differential treatments at all, the whole area being given what seems a reasonable programme for the district. First, as defects in this programme appear, new treatments may be added to try to effect a remedy. Then, as soon as the solution is found, the differential treatments can be discontinued and the remedial treatment applied to all plots. The method has been described more fully elsewhere (Pearce and Taylor, 1948). Usually a 2^n design is best because the questions to be answered are of the form "Is more of this element needed?" not "How much should be applied?" and the trial commonly begins with plots to the number of a power of two arranged in four or eight blocks. The example in Figure XXVI will illustrate the possible development of such an experiment. Initially, it may be supposed, no differential treatments are applied, but after a while deficiency symptoms might appear suggesting the need for more nitrogen. The first diagram accordingly shows half the plots receiving supplementary sulphate of ammonia. Later, the need for more potash might be suggested but, since with some species a high level of potash can induce magnesium deficiency, it might well be thought that two further factors are needed instead of one, and the second diagram shows supplementary sulphate of potash and sulphate of magnesium under test as well as the differential treatment already applied. At a still later stage, it might become apparent that the supplementary nitrogen was indeed essential, and therefore it would be continued on the plots already receiving it and a higher level applied in the others in an endeavour to eliminate the residual effect. Simultaneously, the suggestion might be made that the trees needed either iron or copper, and these two elements might be applied by injection. The results are shown in the third diagram, the brackets round the N, used to indicate application of nitrogen, showing that only the residual effects are here under consideration. At this stage three interactions are confounded, the number of treatment combinations being now four times the number of plots to a block. It is unwise, incidentally, to confound the interaction of all the factors extant at any one time, because this makes difficult the application of further factors. Also, when the number of treatment combinations equals the number of plots, as in the third diagram, some of the high-order interactions must be used as error. After a time the residual effect of the nitrogen applications will have been removed, leaving room for a further factor if one should be suggested, a stage represented by the fourth diagram.

N	N	O	N	O	O	N	O
O	N	N	O	N	N	N	O
N	O	O	N	O	N	O	N
O	O	O	N	N	O	N	O

NKM	NK	KM	N	KM	M	N	O
K	N	NK	O	NK	N	NK	KM
NM	M	M	NKM	K	NM	M	NKM
KM	O	K	NM	NKM	O	NM	K

(N)KMC	(N)K	KMIC	(N)I	KMC	MI	(N)C	I
KIC	(N)IC	(N)KC	C	(N)KIC	(N)	(N)KI	KM
(N)MI	MC	M	(N)KM	K	(N)MC	MIC	(N)KMIC
KMI	O	KI	(N)MIC	(N)KMI	IC	(N)M	KC

KMC	K	KMIC	I	KMC	MI	C	I
KIC	IC	KC	C	KIC	O	KI	KM
MI	MC	M	KM	K	MC	MIC	KMIC
KMI	O	KI	MIC	KMI	IC	M	KC

FIGURE XXVI. Possible successive stages in the evolution of a design with tentative treatments.

In a sense, this approach is a development of factorial design. Many years ago Sir Ronald Fisher said that Nature often replies better to a questionnaire than to a question, meaning that a factorial design with its ramification of interactions often secures more information about the effect of a factor than does a simple one. The difficulty with a long-term factorial trial is that some answers are given at once and the subject need not be pursued further; others are not given at all, although the questions are repeated year after year. If a soil already has sufficient potash and repeated applications of a differential treatment have no effect, the time comes to drop the subject. Perhaps the position may be summed up by saying that

Dame Nature is even more informative in a conversation than with a questionnaire. Sometimes the subject can be pursued, sometimes it should be dropped, either because the lady has answered all that has been asked or because she is just not interested and has nothing to say.

All fertilizer trials need guard rows if they go on for any length of time. Usually single guards will suffice, though double ones may be needed if bulky organic manures are being tested of a kind likely to be pulled long distances by cultivation. There is some evidence that guard rows can be dispensed with if the applications are to be made over only a short period, or if the fertilizer is applied to an area considerably smaller than the space bounded by lines half-way between the trees. Thus, Sharpe and Winsor (1951), working with pecans, fertilized some trees with boron but not others and found no evidence of cross-feeding. Fertilizer was applied in some cases up to 21 feet from the trunks, the trees being at a minimum 50 feet apart. Only one application of boron was made. Again, Lott, Satchell and Hall (1950) gave a single application of radio-active phosphorus to grape vines and reached the conclusion that cross-feeding was negligible between rows, which had been cultivated so as to sever roots near the surface, but that double guards were needed within rows. It is, however, clear from the work of Rogers and Vyvyan (1928, 1934) that, in the case of apples at least, the roots of adjacent trees at commercial planting distances generally overlap to some extent, though this overlapping is limited by competition among the roots themselves, and this means that there must be some cross-feeding. Also, there must be a tendency for roots from a poorly fertilized tree that happen to find themselves in a better fertilized plot to develop unduly and so to accentuate the bad effect of not having guard rows. For coconuts, Salgado (1946) considered guard rows essential.

One general publication of special value is that of Crowther (1947) and Stewart (1947) to be found in the same volume.

61. Trials of Methods of Soil Management including Irrigation

In general, treatments that specify or exaggerate environmental differences are easier to assess than those applied to the trees themselves, because each plant is left to develop according to its own laws and to retain its own internal balances, so that natural relationships, observable between trees treated alike, hold between treatments also. For example, if trunk circumference and tree weight are very closely related, as they often are, then in the comparison of two treatments it would be reasonable to use circumference as an indicator of weight if the treatments differed in respect of the cover crop or the frequency of grass cutting, but not if they were different methods of pruning or tree forming or grafting on different rootstocks, because treatments of the first sort relate to the trees' environment and alter what is naturally variable anyway, while those of the second sort correspond to nothing in the natural environmental variation of the trees and might well disturb internal relationships. In general, the balance of a tree will not be much disturbed by treatments of the kind considered here or by fertilizer treatments either, especially where growth measurements are concerned. With reproductive activity the relationship between measurements is more easily disturbed. For example, irrigation might lead to fruit retention and break the natural relationship between the number of fruit buds and the number of fruit set; no doubt many other exceptions could be suggested. Nevertheless, in general, problems of assessing treatments are at a minimum with trials of the sort considered here.

It is inconvenient to treat individual trees because there are no means of using implements properly in a restricted area, the chief difficulty lying in giving tractors, mowers, etc., easy access to the individual plots as well as room to turn when they are set to work. It is therefore desirable that the plots should extend right across the experimental area from headland to headland or, if that is not possible, from a headland to a landway up the middle of the trial.

Several types of layout can be adopted to secure the kind of plots desired. One is to form plots across the area and to divide them into smaller sub-plots for the comparison of some

other factor. That is perhaps the only case commonly met with where it is reasonable to include two factors together in one experiment even though they are not expected to interact. Large plots have to be used for the first experiment, so the additional factor on the split-plots is studied with no further expenditure for trees and little for additional labour.

Another possibility is to use a criss-cross design (see Section 26), leaving spaces between the rectangles for the turning of implements, supply channels for water or whatever else may be needed.

If long plots cannot be accommodated, the best alternative is to use square ones with guard alleys between them. Treatments of this sort are best compared with guard rows because edge effects can be serious. Where guard alleys are used instead, the treatments should be applied some way beyond the outside trees because of these effects.

Where cover crops are not established when trees are young, calibration is easy to carry out and valuable. Irrigation, however, is usually needed, if at all, from the start, so the chances of calibration are much more reduced. It may well be argued that uniform irrigations followed by a trial of different methods prejudices the issues from the beginning, because the root growth favoured by the initial treatment may handicap the trees with certain of the subsequent treatments. This argument would not apply if the trial were of quantities of water. A similar caution applies to trials of methods of soil cultivation.

The application of herbicides by exponential sprayers has been considered by Brunskill (1957), Dedolph, Basham and Stark (1960) and Dedolph (1960).

62. Spraying and Dusting Trials

Trials of this sort usually need fairly large plots because of the difficulty of confining the treatments within a limited area. Also, a fairly high standard of guarding is called for because pests and diseases can spread as well as treatments. Consequently, it is often advisable to provide single or double guards to keep the treatments to the desired trees and then to record only the middle trees of the plots thus formed. Some relief from excessive guarding can be obtained, however, by deciding on a likely direction of wind at the time of applying treatments and reducing the number of guard rows running in that direction, but this device depends upon reliable weather forecasts. Alleys or screens or both can be used instead of guard trees and are often preferable. Trials of sprinkler methods of irrigation are essentially similar. A general discussion of the problems has been given by Freeman (1967a).

A constant difficulty is the introduction of control plots, which often do no more than declare the obvious while providing a source of infection or infestation for the rest of the trial. It is always wise to include an unsprayed control to show that the pest or disease was indeed present—too many experiments have "demonstrated" that all treatments have been equally successful in controlling an organism that was not in fact present—but it need not be given high replication. Sometimes the pest or disease is likely to increase so rapidly if uncontrolled that unsprayed plots are best disposed to leeward of the main experiment and sprayed commercially if need be to avoid spreading trouble. At other times it may be that unsprayed plots will indeed build up but not to the point that something has to be done about them and it is then useful to have them in the main experiment because experimenters often like to be able to report the level of control they have achieved, but replication need not be high since differences will be large. Designs in supplemented balance are then often useful.

Another treatment with a special status is the one representing current practice. If the others represent suggested improvements, there is a case for giving the standard higher replication since the main comparisons may all involve it, i.e. if the object is not so much to compare the other treatments among themselves as to find out which appear to be genuine improvements and therefore to require further study. Again a design in supplemented balance may be needed. If there is both a standard treatment, which requires higher replication, and an unsprayed control, which requires lower replication, it is sometimes a sound idea to design

the experiment in randomized blocks but to transfer some plots from the unsprayed control to the standard as described in Section 32.

Tentative designs, such as were recommended for use with fertilizer trials (Section 60), are useful here also. A related topic is the repeated use of the same trees. If sprays are applied in one year they may well induce different levels of pest or disease and they may well leave leaf deposits that will affect future levels, but their most lasting effect may be on eggs or spores that will develop and produce differences some time after the previous effects have disappeared on account of leaf drop, migration, etc. In temperate conditions the period is usually one year or less, for in that time the plants will have become dormant, the leaves dropping to the ground and dispersing, thus taking spray deposits with them, but eggs and spores may remain on the bark to give varying levels in the next year. Where that happens a further experiment can be conducted using a row and column design but with original blocks instead of rows and former treatments instead of columns, as explained in Section 33. In a third year it is reasonable to assume that with further dispersal the first set of treatments will have no further residual effect though those of the second set will still need consideration, so a further trial can be designed with the same blocks playing the part of rows and the second set of treatments that of columns with a third set of treatments under study. At each stage it must be reasonable to assume that there is no interaction between the current set of treatments and that of the year before, but that should be all right if spray deposits, etc. have all been dispersed. The statistical designs needed for such experimentation have received a good deal of attention (Pearce and Taylor, 1948; Hoblyn, Pearce and Freeman, 1954; Freeman, 1959).

It is perhaps unnecessary to add that modern spray substances are often highly toxic to human beings, so the greatest care is needed in designing experiments to ensure that those who hold screens, etc. are put to no risk.

63. Trials of Varieties and Rootstocks

In comparing varieties, whether of scions or rootstocks, it will usually be found that the series to be studied falls into groups, such that little interest would attach to a comparison of varieties in different groups, though it is important to select the best within each. Thus, there is no point in comparing a late culinary pear with an early dessert one or a vigorous apple rootstock with a dwarf, but it would be valuable to know which was to be preferred in each class. Hence, some preliminary information should be obtained, either by conducting first a simple observational trial or, in the case of rootstocks, by a histological study of roots, a method which, for many species at least, will indicate the vigour to be expected (Rogers and Beakbane, 1957; McKenzie, 1961; Beakbane, 1967, 1969). With rootstocks a preliminary enquiry is especially valuable if large trees are not to be randomized with small. Some initial information is also needed about incompatible stock–scion combinations so that double working can be used as necessary.

In a preliminary trial of scion varieties it is often helpful to framework (Garner and Walker, 1938; Garner, 1947) on to an existing series of uniform trees. Although such methods are less used nowadays in commercial practice, they retain their usefulness in experimentation as a way of obtaining fruit quickly for assessment of its quality, flavour and appearance. Again, when seedlings are raised there are ways of obtaining fruit quickly (Garner, 1947) without waiting for the new varieties to produce enough wood for grafting and then waiting again until they crop. Other methods exist (Garner, 1947) that reproduce rootstocks quickly so that testing can begin.

Whatever the species, whenever there are large numbers the best way to go to work is to devise a series of "sieves" by which the number under study is whittled down, both by discarding some as obviously unsuitable and by forming those that remain into groups. The approach has been considered in some detail for cocoa by Hoblyn (1945). The proportion of

varieties to get through at any stage must depend upon circumstances—it will obviously be greater when starting with sister seedlings, the result of controlled crossing, than with a diverse collection of oddities sent in by the public—but in general most experimenters appear to pass too many rather than too few. As a matter of fact, a good deal of work has been done on the subject (Yates, 1940; Finney, 1958a, b; Curnow, 1961; see also Finney, 1964) and it deserves more attention than it receives from practical men. It should be noted, however, that most of it was carried out with annuals in mind and consequently with emphasis on crop weight. With most products of perennial species, on the other hand, crude weights are not enough; whether it is pears, tea, hops or grapefruit taste is important, whereas it scarcely enters into the assessment of wheat.

Sometimes the grouping of varieties is obvious, e.g. dessert and culinary, and sometimes it is dictated, at least in part, by the times at which the varieties become available; but often it must to some extent be arbitrary. Thus, suppose a range of rootstocks has been placed in order of expected vigour, then any division into rigidly defined groups must be artificial and, wherever the boundaries are placed, there will be comparisons between groups as important as those within. The method here is to overlap the classes so that some varieties occur in two neighbouring groups.

The difficulty about choosing the number of overlaps is this: if too many are used, the number of treatments to each group will be inflated and precision will fall from a lack of replication. (It is assumed that the total area available for the trial is limited.) If too few are used, the comparisons between groups will suffer. Assuming the boundaries between groups to be entirely arbitrary so that comparisons between varieties in adjacent groups are as important as those within groups, the following rules for the number of overlaps may help. The figure given is in every case a maximum, so fractions should be counted to the whole number below.

(a) To divide a series into two groups, first of all divide the varieties as evenly as possible. Let the mean number of varieties per group be g, then each should contribute $v = \sqrt{(\frac{1}{2}g)}$ to the other.

Thus, if 26 rootstocks, A–Z in that expected order of vigour, are to be divided into two groups, first divide A–M and N–Z, g being therefore 13·0 and v, 2·55. Consequently, two varieties should be contributed by each group to the other to give A–O and L–Z as the final grouping.

(b) To divide into three groups, after the preliminary division, each outside group should contribute v to the inside one, which should contribute $2v$ to each of the outside ones, where $v = \sqrt{(\frac{1}{6}g)}$.

Thus, the 26 rootstocks would first be divided A–I, J–Q, R–Z, making g equal to 8·7 so v equals 1·20. The groups now become A–K, I–R, P–Z.

(c) To divide into four groups, after the preliminary division, each outside group should contribute v to the adjacent inside one, which should contribute $3v$ to the adjacent outside one. Further, the two inside ones should contribute $2v$ to each other, v being equal to $\sqrt{(\frac{1}{12}g)}$.

Thus, the 26 rootstocks could first be divided A–G, H–M, N–S, T–Z, making g equal to 6·5 and v equal to 0·74. Thus the groups become A–I, G–N, M–T, R–Z. This is rather an interesting example, because v, being less than one, has to be approximated upwards to one instead of downwards as usual.

On account of these approximations, it is possible to be left with markedly unequal groups and some adjustment may be needed between their boundaries while keeping the number of overlaps unchanged. Also, in the division into four groups it is as well to keep the number of overlaps the same between all pairs of adjacent groups. Thus here, L could be added to the third group or O to the second. The whole grouping should be done with common sense, the rules providing no more than a guide.

It now remains to make the best possible comparison within each group. Sometimes the whole investigation will be repeated at several sites. That is an excellent development, though it is essential for enough replicates to be used at each site for the results to be studied separately (see Section 13).

In variety trials more than in most other kinds the experimenter should consider carefully how to treat his trees in respect of factors not being investigated in the trial. Thus, characteristics of vigour are less masked by light pruning than by heavy, close spacing too will promote disparities in vigour, while the thinning of fruit will often disguise quality differences.

In comparing varieties it should be borne in mind that with different genetical content, differences in the main characteristics of growth and cropping are only to be expected. The question is not "Do these varieties differ from one another?" but "Knowing they must differ, by how much?" With different scions, it is to be expected that the shoots of the trees will differ in characters genetically determined; but with different rootstocks it does not follow that the shoots will differ, though the roots must. Consequently, rootstock trials may require significance tests, but not trials of scion varieties.

Further, the final choice of a variety is often made on a balance of characters rather than any one outstanding feature. The unwary may here be tempted to use discriminant analysis but that is to mistake the problem. The task is not to find and to give weight to those varietal characteristics that are most distinctive but to those that are of most commercial importance. Thus, a variety with excellent flavour and appearance may be rejected because it travels badly and another selected which, though really excelling in nothing, has no notable defects. Consequently, statistical analysis, though valuable in confirming that a certain character is indeed chiefly associated with a certain variety, does not hold the place it does in other sorts of trial.

The assessment of rootstocks gives rise to much the same problems as the assessment of scion varieties. One point needs special mention. No experimenter comparing varieties would suppose that natural relationships found within trees of the same kind are necessarily going to hold between trees of different genetic content, which often have markedly different habits. Consequently he would not, for example, argue that a variety that gave vigorous growth was necessarily going to have a lot of blossoms in the following year, though the relationship might be well established within varieties. He would therefore record both quantities if he wanted to know about both. The same applies when the same scion variety is grown on a range of rootstocks. The genetical content of the material being studied may indeed be constant, but the rootstocks could lead to differences of habit no less marked than those found between the rootstocks themselves if they were grown without being grafted with a uniform variety.

In raising material for a variety trial it is essential that all trees should have a similar source. It is a mistake to raise some of the varieties on the experimental farm, to buy some from one nurseryman and the rest from another. If the comparison is to be made fairly it is as important to start off with trees raised in the same way as it is to treat them similarly once they are planted. It is also a mistake to plant one variety one day and a second one the next. Admittedly it saves trouble in the nursery though at the expense of trouble in the field. There are two disadvantages. One is the building-in of differences between varieties if there is a change in the weather. The other lies in the successive trampling of the experimental area and the possible induction of compaction patterns.

Special problems arise in seed orchards on account of the need to secure evenness of pollination. They have been considered by Freeman (1967b), Dyson and Freeman (1968) and Freeman (1969).

64. Trials of Pruning and Tree Formation

These call for very little comment. They can mostly be carried out using one-tree plots without guard rows and present few difficulties to the designer, though some experienced

investigators are of the opinion that guarding is desirable because of the wide differences in height and spread that can be induced by pruning, together with the consequent differential competition between the aerial parts of adjacent trees. The effect of shading is also sometimes such as to make guarding necessary. Perhaps the best compromise is to use rather wide spacing between trees, i.e. narrow guard alleys rather than guard rows.

In such trials calibration is sometimes possible, but usually treatments should start as soon as the trees are planted.

One difficulty arises if there is the great variation in shape of tree and types of growth induced by different methods of pruning and tree formation because of the lack of comparative growth records. Thus, peach trees pruned in winter and summer are very difficult to compare in respect of vigour because few shoot measurements made on one set of trees can reasonably be made on the other.

In Section 61 the point was made that changing the environment of a plant leaves it free to develop according to its own internal laws and consequently two quantities that are related in the ordinary way will be similarly related in differences between treatments. However, with the sort of experiments here considered the situation is very different. The treatments are being applied directly to the plants themselves and, what is more, they are of a kind *intended* to disturb the natural relationships and to induce a habit of growth not native to the species or variety. Accordingly measurements can be related only with difficulty to other measurements that would ordinarily tell much the same story. Thus, with apples the increment in trunk circumference and the length of extension shoots usually amount to much the same thing over a period. In any one season, it is true, apical activity can take place only at the expense of cambial and *vice versa*, but each leads eventually to the other and over a period of any length a balance between the two is set up and measurement of one gives a good idea of the other. Under conditions of differential pruning, however, that happy position no longer obtains, because severe pruning continually leads to shoots being forced out at the expense of trunk thickening. Hence, if a full picture of the growth activity of the tree is required, both measurements have to be made independently.

65. Trials of Commercial Systems

Special difficulties attend the assessment of integrated systems of growing a given kind of tree. The variety, the spacing, the farm machinery, the methods of spraying and soil management all have to cohere, a change in one making desirable, if not essential, changes in the others. So factorial a problem defeats factorial design, because the total number of treatment combinations would be so large that no experiment could study them all, while some would be obviously impracticable and others likely to be disastrous. What can be studied are various suggestions for integrating the factors into a consistent system that makes sense to practical men.

Where that is done and complete systems have to be compared, it is clear that detailed records, like shoot growth and bud development, have little place except as ancillaries to suggest why a system is falling short. What is required is an estimate of crop per unit of area, the costs of establishment and maintenance on the same basis and assessments of quality of produce by grades to enable the yield to be related to costs of production. Of course prices will change each year, but that need not enter into the argument initially. It could well be, for example, that one treatment will be best only if early fruit commands a premium. Instead of using actual prices, it is better to clarify the agronomic position regarding yields in various grades and then to assess the economic position using notional prices that represent the market in a range of conditions, e.g. a hypothetical year in which early crop had lost its premium, or one in the future, when processing will be more common. There is no special importance about the market as it happened to be at the time of the trial, as if the prices that prevailed then were of eternal significance.

If economic records are to be obtained, it is no good using small plots that do not admit of effective costing. Although it is no doubt possible to measure the time, etc. spent pruning or picking a single tree, the figure will not in general be representative of commercial practice, while for cultural operations, like preparing the land or cultivating, the use of implements in a small area bears little relation to their use in a large one. It is unreasonable to attempt costing on plots smaller than 0·2 ha and often much larger ones will be needed. Accordingly, replication at a site can rarely be large. Nevertheless, replication and randomization are both essential if standard errors are to be assigned to the means. That applies to extensive trials as to intensive ones and to figures for economic profit or loss as for bud numbers or degree of chlorosis.

That consideration brings up the nature of replication in such an experiment. Basically treatments are replicated when they are repeated in conditions that need not be exactly the same as before but are equally representative of the conditions to which results are to be referred. In the case of commercial systems the differences do not consist only of variability in plants and soil. Much more important are the varying interpretations that different farm managers will put upon the system and their varying success with it, some doing well with this and some with that. In short, each block really needs a different farm with a different man in charge if treatments are to be truly assessed. Further, a research centre can be the worst place for such a block. It is true that the farm supervisor can be asked to do everything to commercial standards, spraying as near as possible to the ideal date but not necessarily exactly on it and picking only what a commercial farm would pick, but all the routines would be against him. An experimental farm pretending to be a commercial one carries little conviction. In the Pastoral Symphony Beethoven has a little fun at the expense of village bands and it is profound music, quite unlike village bands, which nonetheless have a quality of their own not shared with symphony orchestras. The writer has little practical experience to back this assertion because his work has been largely at research institutes, but he believes that the only worthwhile replication in a trial of commercial systems is on a range of commercial farms, each doing its best by its own lights, and the only true error is the interaction of treatments and farms.

66. Spacing Trials

In trials of commercial systems the chief factor is often spacing, the other choices, whether of variety, tree form, hygiene or fertilization, being made in the light of it. However, the opposite problem exists of taking a particular kind of plant and enquiring what spacing would suit it best.

Spacing is not just a question of density. Even if it is decided that there should be 400 trees to a hectare, it does not follow that they must be planted 5 metres square. It may well be that 6 metres by 4·25 metres or 10 metres by 2·5 metres would be better, whether to allow for the passage of farm machinery or for some other practical reason. Indeed, the main object of study could be rectangularity, i.e. the ratio of the distance between rows to the planting distance within them. Even then, there may still be need to consider triangular plantings as well.

There is no objection in principle to comparing spacing treatments using conventional designs in plots and blocks, though they are often unwieldy. For one thing, they call for large plots, because the experimental trees need a complete external guard to preserve spacing, i.e. if there were only one experimental tree, eight guards would be needed, and so for the next plot. In practice much larger plots are used, but even then difficulties are not at an end. Should all plots have the same total area, guards included, or should each have the same area of recorded trees, the guards being additional? In general, the second solution is better, but it does mean that there will be alleys round the plots with different spacing, leading to waste of land and increasing the disturbance of micro-climate, which is a permanent worry with

spacing trials. Thus, a close-planted plot to leeward of a wide-planted one must reduce wind speed and so lead to development atypical of the wider spacing. Most difficulties are much reduced if the row spacings are the same for all treatments; interest then centres on spacing within rows, because then there is little need for guard rows, only for guard trees within rows. Difficulties with micro-climate are, however, no less.

Where conventional methods of design are adopted the advantages of a harmonic series should be noted. If it is proposed to try out row spacings of 4, 5 and 6 metres, a plot has to be 60 metres wide to accommodate respectively 15, 12 and 10 rows. If the spacings are changed from an arithmetic series to a harmonic one, spacings of 4·17, 5·00 and 6·25 metres will allow of plots only 25 metres wide, containing 6, 5 and 4 rows respectively of the three spacings.

Increasingly, however, experimenters are turning to systematic designs for spacing trials, the difficulties with randomized ones being too great. Such designs are of various kinds but most are based on arranging the planting positions in a geometrical pattern that gives either a range of densities at nearly equal rectangularities or a range of rectangularities at nearly equal densities. The rest allow for changes in both characteristics. They have been used for some time (Bocquet, 1953; Freeman, 1964), their recent development stemming from a paper by Nelder (1962). The one situation in which randomized designs hold undisputed sway is that where the distance between rows is the same for all treatments, the object of study being spacing within rows.

To consider first a design intended to study different densities at a square planting, a series of spokes is laid out as in Fig. XXVII(a), all directed to a hub and inclined so that there is a fixed angle, θ, between neighbours. (Experience suggests that a convenient value for θ is about 5 degrees.) A constant distance is then measured along each spoke from the hub to give the first set of planting positions. A second set is found by taking radii so that the new set of planting positions is further out along the spokes by an amount equal to the distance between adjacent trees of the first set. Then a third set is found from the second and so on. Although conceptually very simple the method is unsatisfactory to a conscientious surveyor because errors are cumulative and on irregular ground they can become quite serious.

A more sophisticated and more exact approach is the following, which owes much to Nelder (1962). Suppose it is desired to study a range of densities varying from an area of A_1 per plant to A_N. Rectangularity, R, will be measured as the distance between plants across spokes relative to that along them, so $R = 2$ is agronomically the same as $R = \frac{1}{2}$. If the spacings are to number N, yet to be decided, then on account of the need for guards at each end of a spoke there will have to be $(N+2)$ trees in each, numbered from 0 for the inside guard through $1, 2, \ldots, N$ for the experimental trees to $(N+1)$ for the outside guard. To preserve the radial pattern and to achieve a steady gradation in the areas available to successive trees, each tree will be assigned a space C times as wide and C times as long as that assigned to its inner neighbour. (At this point a departure will be made from Nelder's thinking. Here the boundary between any two trees will be taken as the half-way point between them.) Initially N and C are undetermined. On account of the geometrical progression of successive areas

$$(N-1)\log C = \tfrac{1}{2}\log(A_N/A_1)$$

However, that only establishes a relationship between N and C without determining either. The most important consideration in choosing the number of spacings could well be the maximum length of spoke that can be accommodated, so it is useful to know that the length of spoke L, will be about

$$\frac{2(C^2\sqrt{A_N}-\sqrt{A_1})}{(C^2-1)\sqrt{R}}$$

Also, it is desirable to have enough steps, i.e. N should be at least 5 or 6, and there is no point in considering very small ones, i.e. C should be 1·1 or more. Also, θ should not be too large

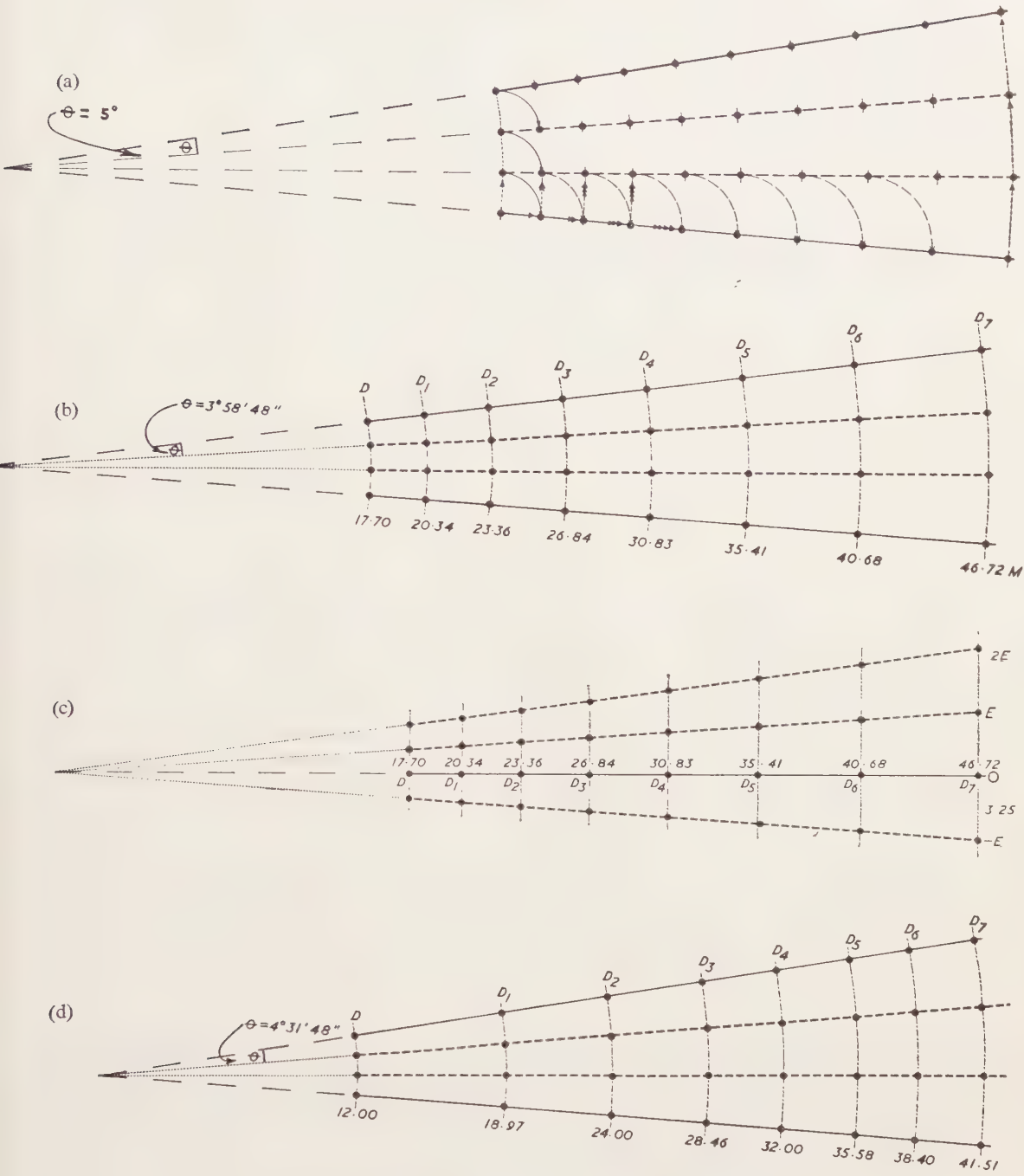

FIGURE XXVII. Examples of different types of systematic spacing trials. (a) is a simple design where all trees are approximately on the square but with differing areas in which to grow, (b) gives a choice of rectangularity and facilitates accurate surveying, while (c) is a modification of (b) with straight alleys for the movement of farm implements. Finally, (d) is a design that gives constant growing areas with differing rectangularities.

and certainly should not exceed 10 degrees because the geometry assumes that all trees have a space that is approximately a rectangle. It is possible to find θ from the expression

$$\theta = 28{\cdot}65R(C^2-1)/C$$

The distance, D, of the inner guard from the hub is found thus:

$$D^2 = \frac{4A_1}{R(C^2-1)^2}$$

From the hub to the first experimental tree will therefore be DC, to the second DC^2 and so on.

Thus, if it were a question of exploring the range of densities from 4 square metres per plant ($A_1 = 4$) to 16 ($A_N = 16$) with a rectangularity of 2 or $\frac{1}{2}$, then

$$(N-1)\log C = \tfrac{1}{2}\log 4 = 0{\cdot}30103$$

and the choice would have to be made from the following table:

		Length of spoke (m)		*θ (degrees)*	
N	C	$R=2$	$R=\frac{1}{2}$	$R=2$	$R=\frac{1}{2}$
5	1·1892	12·5	25·0	19·95	4·99
6	1·1497	14·5	29·0	15·94	3·98
7	1·1233	16·5	32·9	13·35	3·34
8	1·1010	19·0	38·0	11·04	2·76

In such circumstances the choice might well fall upon $N = 6$, $C = 1{\cdot}1497$, $R = \frac{1}{2}$, $\theta = 3{\cdot}98°$, i.e. $D = 17{\cdot}704$ metres. Trees would then be planted along each spoke at the following distances from the hub: (0) 17·70 metres, (1) 20·34 metres, (2) 23·36 metres, (3) 26·84 metres, (4) 30·83 metres, (5) 35·41 metres, (6) 40·68 metres and (7) 46·72 metres. The result is set out in Fig. XXVII(b). Another method is that given by Rogers (1972).

A difficulty with this design is the lack of straight alleys of uniform width in which farm machinery can conveniently operate. A useful modification is shown in Fig. XXVII(c). In essentials the design consists of spokes radiating from the hub and centred on a base line, crossed by parallels, perpendicular to the base line. Trees are planted at all intersections. The geometry has been described by Bleasdale (1967b) and Freyman and Dolman (1971); that given here is similar though not exactly so.

To consider first the case when there are to be an odd number of spokes, the base line will be along the middle one. The values of C, L and D are calculated as before, but θ is different because the spokes are no longer evenly spaced. Instead, a perpendicular is set off at the outermost tree position on the central spoke and points are marked off along it at distances $\pm E$, $\pm 2E$, $\pm 3E$, etc. from the base line, where

$$E = C\sqrt{(RA_N)}$$

These points are then joined to the hub to form the other spokes. Further perpendiculars are drawn from the base line at D, DC, DC^2, etc. from the hub. Thus, in the former example, if $N = 6$, $R = \frac{1}{2}$ then $C = 1{\cdot}1497$, $L = 29{\cdot}0$ metres, $D = 17{\cdot}704$ metres as before and $E = 3{\cdot}25$ metres. The resulting design will preserve areas well along the parallels, though the rectangularity will become rather indefinite towards the sides.

If it were intended to have an even number of spokes the base line would be half-way between the two middle ones and that is less easy for the surveyor. In principle, however, everything is as before except that the points on the outermost perpendicular to the base line are placed at distances $\pm\frac{1}{2}E$, $\pm 1\frac{1}{2}E$, $\pm 2\frac{1}{2}E$, etc. from it, not at 0, $\pm E$, $\pm 2E$, etc.

However, the problem may be quite different. The need could be for a design in which all trees occupy the same area, A, but with rectangularities that range in some even manner

from plots in which the length is R times the breadth to others in which it is R' times $(R' > R \geqslant 1)$. An example is given in Fig. XXVII(d).

It is required then to find distances, $D_0, D_1, D_2, \ldots, D_N, D_{N+1}$ at which trees should be planted along each spoke so that each occupies the same area. To achieve that, the distances must be related in the following manner:

$$D_1(D_2 - D_0) = D_2(D_3 - D_1) = D_3(D_4 - D_2) = \ldots, \text{ etc.}$$

The odd thing about the series is that D_0, D_2, D_4, etc. can be determined without reference to D_1, D_3, D_5, etc. and *vice versa*. There is accordingly an indeterminacy about the design in that D_0, D_2, D_4 can be directly calculated, leaving D_1 and consequently D_3, D_5, etc. at the discretion of the experimenter, who has to face the difficulty that a value of D_1 that gives a smooth series of rectangularities in one part of a spoke may give awkward jumps in another.

A choice will again have to be made between having the larger planting distance across the spokes, in which case the rectangularity at D_1, measured by the ratio of the distance across spokes to that along them, will be R increasing to R' at D_N, or along them, in which case it will be $1/R'$ at D_1 decreasing to $1/R$ at D_N. A further choice that has to be made concerns N, the number of experimental trees in each spoke. As before, it should be at least five and, in general, not more than eight except for some particular reason. However, it must always exceed R'/R.

The calculation begins by calculating

$$V = \frac{R' - R}{NR - R'}$$

which will depend upon the selected value of N. It is recommended that the calculation now proceeds as follows:

$$D_0 = \frac{\sqrt{A}}{\sqrt{R_1}} \frac{1 + \sqrt{(1+2V)}}{V}$$

where R_1 is the rectangularity sought at D_1, i.e. R or $1/R'$. Then

$$D_1 = D_0\sqrt{(1+V)} \quad \text{or thereabouts}$$

$$D_2 = D_0\sqrt{(1+2V)}$$

In general, θ should not exceed 12 degrees. It is found thus:

$$\theta = \frac{57{\cdot}3}{D_1}\sqrt{(AR_1)} \text{ degrees}$$

There does not appear to be a simple, exact expression for L, the length of a spoke, but in general it equals

$$\{\sqrt{[1+(N+1)V]} - 1\}\, D_0$$

or rather more.

Thus, to take an example, if it were proposed to study plots ranging from square $(R = 1)$ to those four times as long as they are broad $(R' = 4)$, all with an area of 9 square metres $(A = 9)$, the following table will help in making the decisions.

First, for $R_1 = \frac{1}{4}$

N	V	D_0	D_1	θ	L (*approx.*)
5	3	7·29	14·58	5·90	24·5
6	$1\frac{1}{2}$	12·00	18·97	4·53	28·7
7	1	16·39	23·18	3·71	32·8
8	$\frac{3}{4}$	20·65	27·32	3·15	36·8

A similar table can be drawn up for $R_1 = 1$, but values of θ will be four times as great and therefore too large to be considered. In practice, $N = 6$ or 7 are the preferred cases and practical reasons, e.g. that L cannot exceed 30 metres, might decide the matter. It will be assumed that the choice has fallen on $R_1 = \frac{1}{4}$, $N = 6$. Accordingly further distances would be calculated thus:

$$\begin{aligned}
D_2 &= D_0\sqrt{(1+2V)} &&= 24{\cdot}00 \text{ metres} \\
D_3 &= (2 - D_0/D_2)\, D_1 &&= 28{\cdot}46 \text{ metres} \\
D_4 &= (2 - D_1/D_3)\, D_2 &&= 32{\cdot}00 \text{ metres} \\
D_5 &= (2 - D_2/D_4)\, D_3 &&= 35{\cdot}58 \text{ metres} \\
D_6 &= (2 - D_3/D_5)\, D_4 &&= 38{\cdot}40 \text{ metres} \\
D_7 &= (2 - D_4/D_6)\, D_5 &&= 41{\cdot}51 \text{ metres} \\
L &= D_7 - D_0 &&= 29{\cdot}51 \text{ metres}
\end{aligned}$$

Some care is needed over this part of the calculation because a slip at one point introduces error into all that follows. A good check is to repeat the calculation using the same values of D_0 and D_2 but a different value of D_1. Values of D_4, D_6, etc. should be unchanged.

The rectangularity at any point varies with the square of the distance from the hub, so the rectangularities in this instance are (1) 0·250, (2) 0·400, (3) 0·563, (4) 0·711, (5) 0·879 and (6) 1·024, the last being equivalent to 0·976. Because there is no simple way of finding L, the rectangularities towards the outside will rarely have exactly the values intended, but the approximation will usually be quite good enough. It is unfortunate that the succession of rectangularities tested is such that values are closer together the nearer they are to one, i.e. the the square case, but that is inherent in the geometry. Another method has been given by Freyman and Dolman (1971).

So far no mention has been made of the number of spokes to a block. Since the outside ones will have to be discarded, it is wasteful to use only a few. Indeed, so long as trees are planted on circular arcs, there is no need to set a limit. If θ divides exactly into 360 degrees a complete wheel is possible. More usually, however, a block (or "fan") is formed of six or eight spokes and repeated. Usually it is as well to have four blocks, pointing in four directions at right angles. That is most important for micro-climatic reasons. A fan that faces the prevailing wind will have a tunnelling effect, while one that faces the midday sun may give results quite different from one that faces the other way. Accordingly, despite its neatness, the complete wheel is not as useful as might at first appear, because for purposes of interpretation it has to be considered in quadrants or other divisions. Sometimes a rectangular piece of land can be used more economically if the wheel is divided, even though guard rows are thereby required. It is obvious that two fans side by side but facing in opposite directions made good use of an area. Whatever happens, infilling the spaces is desirable both for economy of land use and uniformity of micro-climate.

The surveying of a systematic design is made much easier if the hubs are all accessible or, at least, visible. In fact, a hub can be in the forest, on the other side of the laboratory block or in the Director's office, in which cases difficulties multiply. However, it should not be impossible to peg out a central spoke, and to mark planting positions along it with approximate distances between them. The other spokes can be located by measured perpendiculars from the original one. Needless to say, level land is needed but that is not just a requirement of surveying. A spoke that runs downhill will confound spacing with altitude and possibly depth of soil, moisture and exposure as well, so it should be avoided.

Once established, fans are difficult to cultivate evenly and mostly it is advisable to admit defeat from the start as far as machinery is concerned and to use herbicides as much as possible. Mowing grass presents fewer problems, but even the widest spacings have to be mown using machines able to negotiate the closest of planted trees, which can be time consuming.

Fertilization and irrigation raise the problem whether allocation should be at a certain amount per tree or per area, while pruning can be either uniform or adapted to the space available for the tree to grow in. Such decisions have to be based on commercial practice with different spacings, unless the intention is to discern the effect of spacing *per se*. In that case fertilization and irrigation should be uniform on an area basis and trees should ideally not be pruned. Control of pests and diseases presents similar problems but usually it is best to secure good control even at the cost of spraying more intensively on the closer planted trees, but even here there is a case for uniform treatment, e.g. if it were a question of seeing if insects did multiply more in the greater shelter given by closer planting.

Pollination also raises problems. Since varieties that are not self-fertile are necessarily grown with a pollinating variety anyway, it is not clear what objection need be raised to having alternate spokes of two different varieties. If that is felt to be wrong, the only possibilities appear to be the introduction of potted pollinators at critical periods and pollination by hand.

The analysis of data must be carried out by regression methods. Initially each block or fan should be studied separately and then tests should be made to see how far the regression relationships found are common to all. If relationships are different in different blocks, it is helpful if a good reason can be suggested.

Where trees are planted on arcs with a constant characteristic, whether of density of planting or rectangularity, nothing is gained by considering trees individually; it is enough to analyse the total or mean of all comparable trees within a fan. Where, however, trees are not on arcs each must be considered as an individual. Usually it is better to work not with a measure of crop or growth but with its reciprocal, i.e. $1/W$ instead of W (Holliday, 1960). Similarly, instead of plot dimensions, it is better to work with plant densities, i.e. the number to a unit area. More elaborate relationships have been considered by Bleasdale and Nelder (1960), Bleasdale (1967a, b) and Farazdagh and Harris (1968). General studies of the relationship between development and spacing of plants have been made by Mead (1966, 1967) and Berry (1967).

Where the independent variate is to be a measure of shape the position is less clear. The activity of the tree, crop weight or growth can be used directly or as its reciprocal. Similarly, the rectangularity can be measured from one upwards or from one to zero. Thus, for the last example rectangularities were 0·25, 0·40, etc., and that is probably better than 4·0, 2·5, etc., though the meaning is the same, because it avoids extreme values, which are always a nuisance in regression analysis. If rectangularity is to be measured in that way, it is probably better to measure the activity of the tree directly since the resulting measure is more likely to be within similar finite limits. (It should be recalled that the reciprocal of zero is infinity and the result of extreme close planting could be zero crop or zero growth.)

67. Trials of the Spread of Diseases

In studying a disease it is often valuable to be able to detect patterns in its manner of spread. Starting with a single diseased plant the symptoms may spread in all directions or down the wind or in the direction of cultivation. If any such tendency can be established, useful suggestions follow how the pathogen travels and therefore how the disease might be controlled. It is as well to heed the warning of Van der Planck (1963) that diseases usually spread by a series of leaps that create new foci of infection

In general, the grading of symptoms does not help much. It is better to work with dichotomies; either a plant does show symptoms or it does not; either it is dead or it is alive. Gaps present a difficulty if it is not clear why the plant died. Was it a premature victim of the disease or did it die from other causes? For that reason incomplete plantings can be a nuisance. Also, the most informative results are often found when about half the plants or rather fewer are infected. At the two extremes, i.e. when all are diseased or none are, there is obviously no information to be gained.

Investigations of this sort are broadly of two kinds, temporal and spatial. In the first a clean area is taken and some diseased plants are introduced into it. Here there are advantages in a triangular planting because all neighbours are equally spaced; each diseased plant is surrounded by a hexagon of clean ones, which in turn is surrounded by a further hexagon of twelve other clean ones and so on. After any recording a list can be made of clean plants classified according to distance and direction from a focus of infection. At the next recording the proportion of plants in each class to have succumbed to the disease can be noted. If significance levels are required either the χ^2-test or the binomial theroem may be used. Incidentally, the frequent recording of an area may itself disturb the phenomenon being investigated, *Homo sapiens* being an effective carrier of some pathogens, especially of soil fungi.

In spatial investigations the experimenter is called to an area that exhibits an outbreak and he is trying to discern the manner of spread from what he sees at that particular moment. If there are gaps, or "vacancies", he might be in an awkward position because he will often not know whether they were caused by the disease or arose for some different reason. From the point of view of assessing the situation the distinction could be vital.

However, he can do no more than map the spatial incidence of disease at that particular moment. He may indeed attempt a grading, though he would be well advised to attempt a separate analysis for each level of severity rather than to amalgamate grades into an overall measure. He then has to study his mapping to see if there is any discernible pattern.

The most useful statistical study remains that of Freeman (1953), who also gave a full bibliography. His method relates to a rectangular area, $m \times n$, with square planting. It is true that fields are rarely exact rectangles but it is not necessary to record everything and an area of m rows and n plants in each row is often convenient. Freeman distinguished two cases, the "queen" case, in which the disease, like a queen in chess, may travel equally down rows, across them or along the diagonals, and the "rook" case in which it travels down or across rows only. As he remarks, the case in which it travels only along rows is the rook case with $m = 1$.

To take an example of his method, there will be $b\ (= mn)$ planting positions, of which four will be in corners, $(2a-8)$ along borders, where $a = m+n$, while $(b-2a+4)$ positions will be interior. Since, in the queen case, each corner position has three neighbours, each border one has five and each interior one has eight, there are A "joins" of neighbours where

$$2A = 8b-6a+4$$

The doubling of A reflects the fact that each join has been counted twice, once from the point of view of each of its end positions.

If there were no vacancies that figure could be accepted as it stands. Suppose, however, that u of the interior positions are vacant, as are v of the border ones and w of the corners, and write $u+v+w = U$. Suppose also that p of these vacancies adjoin one another, then the true value of A is

$$4b-3a-8u-5v-3w+p+2$$

while the number of plants is $b-U$.

If now n_1 of the plants exhibit a certain condition, it is to be expected that the number of joins that connect two such plants will be

$$\frac{An_1(n_1-1)}{(b-U)(b-U-1)} = \mathrm{E}, \quad \text{say.}$$

A count can be made and could well be found to agree with the expected figure. That implies a random distribution of the condition. It might, however, be larger than theory, thus indicating some form of contagion, while in some instances it might be less.

In order to judge whether apparent departures from randomness are significant, it is necessary to take the argument further and to enquire how many groups of three exist. At this point some care in definition is required. Suppose that there were six neighbouring plants, thus:

A B C
D E F

and suppose that A, B and F all showed the condition under study, then there would be one "dependent" join, i.e. ordinarily two joins would involve four plants, but these involve only three. Suppose, however, that A, B and E showed the condition, there are three dependent joins, ABE, BEA, EAB, according to which of the three plants is to be accounted the middle one. In order to find the number of dependent joins, B, it is easiest to count $x_1, x_2, x_3, \ldots, x_8$, which are respectively the number of existent interior plants to adjoin $1, 2, 3, \ldots, 8$ vacancies. Similar counts are needed of $y_1, y_2, \ldots, y_5$ the numbers of existent border points to adjoin $1, 2, \ldots, 5$ vacancies and z_1, z_2, z_3, the numbers of existent corner points to adjoin $1, 2, 3$ vacancies. Then

$$B = 28b - 36a - 28u + 10v - 3w - 7x_1 - 13x_2 - 18x_3 - 22x_4 - 25x_5 - 27x_6 - 28x_7$$
$$- 28x_8 - 4y_1 - 7y_2 - 9y_3 - 10y_4 - 10y_5 - 2z_1 - 3z_2 - 3z_3 - 44$$

Also, the variance of the actual counts of joints about the expected value, E, is

$$\frac{2B(n_1-1)(n_1-2)n_1}{(b-U)(b-U-1)(b-U-2)} + E - E^2 + \frac{[A(A-1)-2B]n_1(n_1-1)(n_1-2)(n_1-3)}{(b-U)(b-U-1)(b-U-2)(b-U-3)}$$

The method can be extended to the case where two conditions are observed, provided they are mutually exclusive. The second count, leading to n_2, might be of healthy plants, i.e. those *not* having the condition, in which case $n_1 + n_2 = b - U$, but it might be of recent death from an unknown cause that was suspected of being the culmination of the symptoms observed in the first count, or it might relate to other symptoms that could be an alternative expression of the disease. It is important only that no plant can be included in both counts. In that case, given a random distribution of either or both conditions, the number of joins between the different conditions has the expected value of

$$\frac{2An_1n_2}{(b-U)(b-U-1)} = E', \quad \text{say}$$

with a variance of

$$\frac{2Bn_1n_2(n_1+n_2-2)}{(b-U)(b-U-1)(b-U-2)} + E' - E'^2 + \frac{4[A(A-1)-2B]n_1n_2(n_1-1)(n_2-1)}{(b-U)(b-U-1)(b-U-2)(b-U-3)}$$

The rook case is similar except that

$$A = 2b - a - 4u - 3v - 2w + p$$
$$B = 6b - 6a - 6u - 3v - 2w - 3x_1 - 5x_2 - 6x_3 - 6x_4 - 2y_1 - 3y_2 - 3y_3 - z_1 - z_2 + 4$$

It should be noted that only joins along or across rows are now being considered, so p is not necessarily the same as before, nor are x_1, x_2, x_3, x_4, y_1, y_2, y_3, z_1 and z_2, while x_5, x_6, x_7, x_8, y_4, y_5 and z_3 are necessarily zero because there are not so many joins emanating from any point.

If it is a question of considering only spread along rows, i.e. the linear case, there is no difficulty about counting the number of existent joins in any row and adding the figures for each of the n rows, so producing A. The number of dependent joins, B, is likewise readily

found, being equal to the number of plants that show the condition and have two neighbours that show it too.

Freeman emphasized the need for the temporal study of any situation and he is clearly right. The best way to study the spread of a disease is to watch it spreading. Nevertheless, there are times when an investigator faced with a new situation has no time to spare; he has to look at the spatial pattern before him and infer from it as much as he can. It is then that these methods have their value. Freeman suggests that the linear case be considered first. It can be applied along rows and also across them, if the latter direction corresponds to any farm operation. If the distribution in the linear case proves to be non-random, the rook case calls for consideration and probably the queen case as well.

Other methods that may be of interest are given by Van der Planck (1963), especially in his concluding chapters.

68. Trials with Several Kinds of Treatment

The use of factorial designs has led to the question how far several different trials should be combined into one big one. The common-sense answer would seem to be that where there is a reasonable expectation of interactions appearing, the factorial design is desirable.

Trials of new varieties are usually better not combined with other experiments. It is admittedly tempting to design, say, a manurial trial with large plots split into one-tree sub-plots for varieties. Although it is admirable for the new varieties, which are thus tested under a range of manurial conditions, it is disastrous for the manurial trial. If this had been carried out on known varieties, selected to give a wide range of characteristics, it would be possible *either* to assert, in the absence of an interaction, that a certain recommendation can be made of all varieties, *or*, if there is one, that it can be made for certain important ones. With new varieties the range may prove too limited to base any generalizations upon it, while an interaction cannot be given the same practical applications. No one wants to know that Seedling 999 is improved by high nitrogen though not Seedling 1001, but which varieties actually being grown may be expected to benefit. On the other hand, new varieties may be combined with another factor in special cases. Some species, for example, show an interaction between variety and method of propagation.

On the other hand, once something is known about a new variety any further test must be as part of a system. If the existing standard variety has a spreading habit and is resistant to fungal diseases while the new one is more upright and is suspected of needing more thorough spraying, the only realistic way of comparing them is as two systems, each based on one variety, with all concomitant differences of spacing, pruning, spraying, etc. included with the varietal difference.

Nevertheless, it is often a good plan to carry out an investigation using a wide range of standard varieties. That is especially true of rootstock trials and of experiments intended to study methods of pruning and of tree formation. It should be emphasized that this is an application of a general rule that subsidiary factors introduced to extend the applicability of results for the main factor should be *representative* rather than *experimental.*

Although cover crop trials often have a manurial factor added, the interpretation can be rather difficult. Thus, nitrogen added to a grass and clover mixture can kill out the clover, though with low nitrogen the clover may thrive at the expense of the grass. In fact, the treatments do not remain what they purport to be and the resulting interaction can be very complicated.

CHAPTER 7

THE ANALYSIS OF RESULTS

". . . to calculate is not in itself to analyse."
EDGAR ALLAN POE, *The Murders in the rue Morgue.*

70. Some General Considerations

An experimenter with a body of data is usually well advised to study it carefully before starting analysis. Quite a simple summary may show that the treatment differences are so small as not to be worth bothering about. Of course, in trials of a fundamental nature the fact that there is an effect at all may be important in itself, but usually a difference matters only if it is large.

The same preliminary study may also suggest which periods in the life of the trial are most worthy of attention. Thus, after the application of herbicide differences may arise that disappear later and consequently an analysis, to be of use, must be confined to the earlier data. Such concentration of attention is useful, provided it is done objectively and without any tendency to select for study those years in which the results confirm the experimenter's expectations.

Also, preliminary study may suggest conclusions that were not in the experimenter's mind when the trial was initiated, sometimes valuable ones. It must, however, be remembered that no data can confirm conclusions that they themselves suggested, so further evidence has to be sought before anything can be regarded as proved.

Once the preliminary investigation has been completed, the analysis of data can begin. It is the final step in the long journey that began with the formulation of a problem and the inception of an experiment to solve it. It has gone on through all the stages of that experiment, the selection and planting of the trees, the application of treatments, the continual surveillance and the recording of data. All should have been subordinated to the object of the journey, namely, the solution of the problem. The final stage, as Huet (1961) has justly remarked, is at once the easiest and the most dangerous. It is easy because the computer will do the work, whether guided by intelligent purpose or not the experimenter must decide. It is dangerous because inept or ill-considered analysis can throw away information that years of labour have accumulated unseen. The analysis of data is like reading the surface of running water. The experienced can detect movements and irregularities that betoken submerged rocks and the presence of fish, while the inexperienced and the unheeding see only what breaks the surface.

Before analysing data there should be no doubt what is being investigated. In any test, even if applied to a set of random numbers, there is one chance in twenty of a significant result being obtained by chance, so, if enough tests are carried out, sooner or later something will be found significant and that something may well be spurious. The method of analysis to be adopted should be uniquely designated at the time the experiment is designed and should not be departed from. Even the preliminary examination of data, already recommended, should not lead to modifications in the designated series of tests except to suggest the omission of certain of them as needless. In any trial, the analysis of variance enables the experimenter to judge the accuracy of his results and thus to form an opinion about the adequacy of his techniques and the reliability of the conclusions indicated by the data. Sometimes that is the sole reason for analysis; sometimes it is subsidiary to the need for significance tests. The purpose of those tests is to give an answer "Yes" or "No" to questions that were in the experimenter's mind when he was designing the trial. They should not be used to answer questions posed later after the data have been examined.

The only exception to this rule occurs in a series of similar trials. Here, if an effect draws attention to itself in an early trial, it may be tested for in later ones. With annual species the

exception is a large one because repeated experiments are customary, but with perennials a decision has usually to be reached on the basis of one experiment, although there may be several as, for example, when fungicides are tested in a number of different seasons.

A further decision initially required relates to the composition of error. Basically the experimental error is the interaction of the treatments and the various circumstances, i.e. times or conditions, in which they are compared. However, if it is supposed that inconsistencies arise solely from random errors, it is permissible to supplement the interaction with other measures of those random effects, as is often done with extended designs (see Section 30) when the variation between similarly treated plots in the same block can be accounted as error. In a trial with many factors there are interactions of high order in which it is very difficult to believe. (To take an example, if $A \times B \times C$ is confounded, then *a fortiori* higher-order interactions that comprise it, like $A \times B \times C \times D$, cannot seriously be under study.) Such interactions can reasonably be merged with error. Indeed, in singly or partially replicated experiments there is no other way of forming an error line. There is no objection to this practice provided the interactions that are to make up the error are chosen beforehand. It would be quite wrong to work them all out and to choose the smallest.

It should scarcely be necessary to mention that the unit of a field trial is the plot (or, in split-plot designs, the sub-plot) and therefore data for analysis should first be expressed as either plot totals or plot means. Where there are several trees to a plot, each measured separately, the individual figures are not needed.

When the analysis is finished some discretion is called for in its interpretation, especially in respect of significance levels. It is a great mistake to suppose that treatment differences are either in the black shade of non-significance or the bright light of significance, because really there is a gradation. However, as has been explained in Section 12, significance is not necessarily important and care should be taken to answer the questions in the experimenter's mind. The tradition by which editors require an author to say if his results are "significant" or not is most unfortunate and it should be resisted whenever it is unappropriate, e.g. in instances when the object *is* to estimate the magnitude of a difference, the existence of which is undisputed.

All these considerations require thought apart from the calculations themselves. Here the electronic computer has led to a radical change. In the past there were people who were accounted to "understand statistics" because they knew the form of the calculations (in technical language, the "algorithm") that transformed a body of data into an analysis of variance. Nowadays the algorithm is effected by the computer and anyone who claims to "understand statistics" must prove it by showing that he is able to make good use of the analysis produced, not by the unnecessary toil of producing it. Clarke (1965) has emphasized the importance of a good account of what has been achieved.

The analysis of variance having been completed, it is important to study its various parts in a considered order lest an impression gained from one line be questioned by another. For example, the error line shows how well the experiment has been conducted and ordinarily it would be looked at first. If the variance is larger than was expected it may still be possible to do something about it, not by changing anything in the field because the fault must lie in the past, but possibly by recording something that should have been allowed for and calculating a covariance adjustment or by a more detailed recording of the quantity under study. Also, if the error variance is large the apparent non-significance of some expected effects will be more readily accepted. Having looked at the error line, the experimenter will turn his attention in a factorial trial to the highest-order interaction, not to the main effects, and will pass progressively to simpler ones until something requires interpretation. Again, with a gradation of fertilizer applications or intensity of spraying, it is important to study the lines of an analysis of variance in a reasonable order. Suppose that there are three levels of a fertilizer element and the two degrees of freedom between them have been partitioned into a linear and a parabolic effect, it is the latter that should be examined first. If the relationship is in fact curvilinear, the slope is not constant and the estimation of a general value for it, as given by

the linear component, is of doubtful meaning. Conventions vary but most computer programs put error at the end, main effects above interactions and the linear effect above the parabolic, so much of the time the output has to be read from the bottom upwards.

Computers sometimes have to be made to repeat analyses in a different form and programs should be flexible enough to allow for all the possibilities. Thus, suppose that there had been three tree-forming methods tried out on four varieties A, B, C and D. The initial analysis would partition the eleven degrees of freedom thus:

Tree-forming	2
Varieties	3
Interaction	6

However, once it becomes clear that there is an interaction, it is no longer reasonable to study main effects, so it could be wise to ask for the partition:

Tree-forming with Variety A	2
Tree-forming with Variety B	2
Tree-forming with Variety C	2
Tree-forming with Variety D	2
Varieties (as before)	3

Many similar instances arise. It cannot be emphasized too strongly that there are no mechanical rules in the interpretation of data. The point has been made that the form of analysis should be determined at the inception of the experiment. It is wrong to partition and repartition until something appears. However, that should not preclude an ordered sequence of partitions, such as the one just described, which proceeds from the logic of the experiment and could have been foreseen from the beginning, at least as a contingency.

Special caution is needed before modifying the error line. A decision should have been made at the start about what exactly is to be kept separate as something to be studied and what is to be accounted as error. It is a mistake to separate out the high-order interactions and to merge each with error if it is small and to keep it separate if it is large. Such a procedure can only lead to under-estimation of the residual variation. Of course, there will be times when the experimenter is faced with an interaction so large that some notice has to be taken of it even though it is completely unexpected. If he does decide that he must keep it out of error, he should be consistent and make all the concomitant changes, e.g. in not looking at main effects and interactions contained in it until he has looked at the multi-way table of means, as described above. He should not just keep it out of error without taking it seriously as part of the treatment effects.

71. The Analysis of a Single Experiment

With perennial species it is not usually practicable to carry out an experiment more than once. It is true that several years' results can be obtained from the same trees but that is not the same thing, because successive sets of data are not independent and it does not prove much if they do confirm one another. With no means of checking doubtful conclusions, probability statements, such as those given by significance tests, have to be used with caution and consistency.

Above all it is necessary to be clear what effects are under investigation. Thus, if several factors are included together in order to study their two-factor interactions, there may well be a host of higher-order interactions that arise from the experimental design rather than from any concern to investigate them. If there are a lot of them and each is tested separately instead of being merged in error, some are almost bound to appear significant since each has one chance in twenty of doing so by pure hazard. If, on the other hand, some comprehensive

test is carried out first, there is the risk of a significant effect being lost in a swarm of non-significant ones, the whole being non-significant. With repeated trials, if one of these interactions draws attention to itself, it can be looked for on later occasions, but what if there is only one trial?

A similar difficulty arises when the levels of a factor form natural groups. Thus, if an experimenter is studying the action of six fungicides, three derived from copper compounds and three from sulphur, it might well be that there are no differences within either group but a marked difference between them. This last could easily be missed if the five degrees of freedom between fungicides were tested together, the one degree of freedom associated with the difference between groups not being separated from the four degrees of freedom associated with the non-significant differences within the two groups. Again, the position is easier if the trial is going to be repeated, the difference between the groups appearing every time, though the whole treatment mean-square is never significant.

There does not appear to be any completely satisfactory answer to these questions, though several have been advanced. The recommendation here is to eschew mathematics and to return to the objectives of the experiment. An electorate required to appoint a committee or a parliament does not disperse its votes over all who are eligible for election, but first goes through a nomination procedure, whereby those whose claim to appointment justifies serious consideration are distinguished. Voting then takes place for them. Not all who are nominated will be elected; also, some constitutions provide for "write-in" votes whereby it is possible for someone to be elected who is not a nominee. Nevertheless, the practice reduces a procedure that could be chaotic into something like order. In similar manner, confronted with a list of interactions it is possible to nominate those that are actively under study, whereas others are present only because the factorial design produces them.

In general, it is reasonable to regard each main effect as nominated (or why was the factor included at all?) and also each two-factor interaction (or why were the two factors included together in one trial?). Higher-order interactions are usually unnominated, but exceptions may be made either way. Thus, in a pruning trial it might be wise to include a wide range of rootstocks in order to compare the pruning treatments on trees of diverse vigour, but there would be no point in testing the main effect of rootstocks when the whole point of having them is their known diversity. Indeed, if the rootstocks were planted as the main plots of a split-plot design their comparison might be so insensitive that the main effect did not appear to be significant, but everyone would go on taking it for granted. On the other hand, there could well be three-factor interactions of the highest interest and they would naturally be tested. Before testing each nominated effect separately, the unnominated effects can be tested together by adding their degrees of freedom and their sums of squared deviations and carrying out an *F*-test on them all. Only if this comprehensive test gives a significant result are the unnominated effects studied separately to find out which is the operative one. In this way the chance of being misled by a random result is kept within bounds. Where there are only three factors, the last order interaction automatically acquires the status of a nominated effect because it is the only one in the comprehensive test. In split-plot designs the comprehensive test is made difficult if the unnominated effects are spread over two or more analyses and thus have different errors. A test does exist (Cochran, 1951) but is not of a sort to be undertaken lightly.

In the other case, where the levels of a factor form natural groups, it is necessary to nominate the groups right at the start. Then tests can be carried out to see if there are differences *within* any of the groups. If one of them does appear to contain diverse components, it is split according to the differences found. Finally, testing takes place *between* groups.

Others may prefer systems of their own devising and certainly practice varies a great deal. What is important is that testing should proceed according to a regular plan, drawn up when the trial was designed to answer the questions being asked. That is not to say a research worker should be uncritical. If A interacts with B, B with C and C with A, most people would

want to know something about the three-factor interaction—but where there is no chance of securing confirmation of doubtful results the experimenter must come down on one side of the fence or the other and should act so as to be able to rebut any accusation of caprice. In particular, he must not partition and repartition until he finds something. With sufficient ingenuity and determination it is possible to obtain results in this way from almost any trial; the right way is to think what questions are being asked and how they can most sensibly be answered.

Although the analysis of results must not be carried out arbitrarily but in a regular manner, care must be taken to adapt the procedure to the questions being asked. For example, a horticulturist may know that an inorganic compound fertilizer gives beneficial results, but be in doubt whether that is the result of the nitrogen or the potash it is known to contain. Consequently he designs a factorial trial to find out and it might well be identical with that of another horticulturist, who is enquiring if nitrogen and potash bring about any improvement. Despite the identity of designs, the method of interpretation will not be the same, for with one investigator it is an article of faith that there must be a real effect somewhere, while the second is ready to believe that no addition of either element is called for. Thus, if no effects prove significant, the first would still accept as real the one that came nearest to significance. (The position would be most unsatisfactory for him and further investigation would certainly seem called for, but his choice would be quite reasonable.) It is not unusual for past knowledge to influence the interpretation of results.

These are by no means the only examples that could be given. To sum up, an experimenter with no opportunity to check doubtful results is best guided by some standard procedure, though he should be careful to distinguish the cases to which it will not apply. It should be added that the interpretation of results is still a matter of insight even where the significance levels are objectively determined.

72. Significant Differences and some Alternatives

It is an old practice to take the standard error of a difference and to derive a "significant difference", or more precisely a "least significant difference", from it. All that is required is to multiply the standard error by the quantity t taken from tables (Fisher and Yates, 1938) according to the significance level required and the number of degrees of freedom for error. So general has this practice become that results can scarcely be presented without significant differences being quoted and some editors will not accept a paper that lacks them. That is a great pity because the wonderful quality of the analysis of variance is its flexibility, but all that is lost if results may be discussed only in one way.

That is not to say that significant differences are without value. At the least they provide a convenient way of presenting a t-test and one that may be used whenever that test is valid. For example, given only two treatments there is no other test available and that is so whenever an F-test would have only one degree of freedom for the effect under study, e.g. in tests of linear and parabolic effects or of correlation coefficients it is always possible to express the test as a value needed for significance with which actual values can be compared.

However, when there are three or more treatments its use can be more problematical, but even then it may be completely valid. Thus, it could be that a number of new treatments are each to be compared with a standard, or, as in the example in Appendix III, a number of treatments are each to be compared with an untreated control. In such cases it is convenient to be able to state that the standard or the control gave a certain mean value and that any treatment with a mean that differs from it by more than a certain amount can be considered to show a significant difference from it. A pedant may object that the various t-tests thus performed are not independent. Nor are they, because all comparisons involve the same estimate of the mean of the standard or the control. If it comes out too high or too low all comparisons will be affected in the same way and to the same extent, but no one has required that the tests should be independent. The pedant could be satisfied by obtaining separate

estimates of the common value, one for each of the other treatments, but, if that were done, a sensible person would merge them to obtain a better estimate for general use. Thus, dependence would be restored.

Sometimes a more complex situation arises. For example, an early spray of a substance, *A*, is to be compared with an early spray of another substance, *B*, and also with a later spray of *A*. It is natural to ask why the factoral design was not completed with a later spray of *B*, but there could have been serious practical objections, e.g. that a spray of *B* near blossom time would do damage. In such a case the two *t*-tests, however carried out, might well be best despite their lack of independence. If it could be said from the start that the object was (a) to compare early and late spraying using *A* and (b) to compare *A* and *B* as an early spray, no one should object to the testing procedure. The trouble will come if it appears, for example, that *Early A* is rather but not significantly better than *Early B* and rather but not significantly worse than *Late A* and someone wants to compare *Late A* and *Early B*. As a lead to a further trial the proposal is unobjectionable. After all, how a man gets his ideas is of no importance so long as he obtains experimental confirmation before he announces that they work. It could well be that success in spraying depends upon getting both the substance and the timing right and the two effects under study are not going to tell the whole story. Nevertheless, if the experiment had been conceived as two, admittedly dependent, tests of timing and substance it should be kept as such. A hypothesis can never be proved by the data that suggested it in the first place. The difference between *Late A* and *Early B* was not nominated and, though it may be tested, it can be established only by further data.

However, where there are several treatments and no special comparisons justify nomination, it is usually safest to put all degrees of freedom together and to apply an *F*-test. If the result is positive, the most reasonable assumption about the means is that they are all different, e.g. if there is doubt whether rootstocks can affect the time of blossoming of the scion, an *F*-test may show that they can, but in that case the most reasonable conclusion is that each has its own characteristic effect. In general, there is no reason why the treatments should form groups, each uniform within itself. If anyone replies that the groups obviously exist, it is a pity that they were not nominated, for in that case the comprehensive test, with all effects merged, would not have been applied but some more sensitive approach used instead.

Where there is no reason for expecting treatments to form groups, some reservation must be expressed about methods for grouping them. Such methods are the use of the various "multiple range tests" (Tukey, 1949a; Keuls, 1952; Duncan, 1955; Scheffé, 1959) which place the treatments in groups such that those within a group may be considered not to differ significantly from one another. The problem might be realistic if the property measured depended upon only a few genes, but applied to crop or growth the techniques have an air of unreality.

It should, however, be noted that, whatever test is used, it is always possible for two treatments to be different in intention but virtually identical in action. Thus, two annual cover crops may be quite different botanically but, if both germinate too late, they may be equally ineffective. Again, two sets of pruning instructions may look different, but both may lead in practice to the same wood being removed from the trees. An experimenter confronted with two treatments that appear to have much the same effect may well ask himself if they are really as different as they appear to be.

In split-plot designs the computation of significant differences needs a little care. If a factor X has been applied to main plots the significant difference between means representing the main effect of X should be calculated using the error mean square and error degrees of freedom for the main plot analysis, i.e. the first analysis. Equally, if a factor Y has been applied to sub-plots the significant difference for the main effect of Y should be derived from the error of the sub-plot analysis, i.e. the second, and so on. The difficulty comes with the interaction, $X \times Y$. For the comparison of two means representing treatment combinations having the same level of X but different levels of Y, it is correct to use the error for the sub-plot analysis. If the treatment combinations differ in X, whether they are or are not the same

in respect of Y, for a design in randomized blocks the error may be obtained thus (Taylor, 1950).

The appropriate error mean square is

$$\frac{s_1^2+(L-1)s_2^2}{L}$$

where L is the number of levels of Y, s_1^2 is the main plot error mean square and s_2^2 is the sub-plot error mean square. To find the number of degrees of freedom, F (which is not in general going to be a whole number), it is first necessary to work out c, which equals

$$\frac{s_1^2}{s_1^2+(L-1)s_2^2}$$

and then F is given by the formula,

$$\frac{1}{F}=\frac{c^2}{f_1}+\frac{(1-c^2)}{f_2}$$

where f_1 and f_2 are respectively the degrees of freedom of s_1^2 and s_2^2. The appropriate expressions in other cases can usually be derived from the work of Satterthwaite (1946) or that of Welch (1947, 1949) and Aspin (1948, 1949).

73. Heterogeneous Errors and Transformations

In the analysis of variance it is assumed that the variability of all data from their expected values is the same irrespective of treatments, but often that is not justified. For one thing a treatment that involves considerable manipulation of the plant material, e.g. in pruning or potting, may well introduce variation that does not arise in the untreated controls. For another, the mere fact of moving a treatment mean along the range of possible values may lead to a change in variability, e.g. if at a certain temperature no seeds germinate there is zero sampling variation, but at a higher temperature, if half the seeds germinate, the sampling variation is at a maximum, only to decline again when germination is complete. The consequences can be serious. Thus, an estimated error is obtained from all the data and is then applied to all means. If some are more variable than average, differences between them will be judged significant too easily; while for those that are less variable, detectable differences will be spuriously difficult to establish. In extreme cases, the results can be absurd, but even more dangerous are the less extreme ones in which silly conclusions present themselves and are accepted.

The seriousness depends upon the test procedure adopted. An overall test, like the use of F, may be little affected, but a test specific to particular means, like a least significant difference or a multiple range test, is especially vulnerable. The position is no less objectionable when estimation is involved. All means will be assigned the same standard error even though some are obviously better determined than others. Again the extreme case will draw attention to itself, perhaps because putting limits round one of the smaller means will include a long sector of negative values, even though the quantity estimated is inherently positive.

If there are enough degrees of freedom for error there is a simple remedy. Any degree of freedom for treatments implies a contrast. If that contrast is worked out for each block separately it is a simple matter to enquire if the mean of the values obtained differs from zero. Thus, given b determinations of a contrast, one from each block, the total of the determinations squared and divided by b gives the sum of squared deviations for the general effect of the contrast and has one degree of freedom. The sum of squares of the b determinations *less* the quantity already obtained represents the corresponding error with $(b-1)$ degrees of freedom. An F-test is available. However, the method (see Cochran, 1938, and the example by Wadleigh and Thorp, 1940) can be used only with block designs, because it depends upon grouping the

data to provide from each block a separate and independent value of the contrast or other quantity under study. Where there are no blocks there is no way of grouping the data, while in a row-and-column design the grouping is not unique. Also, without very adequate replication the method does not give enough degrees of freedom in the separate errors to be of much use. Nevertheless, where it can be used, by providing each quantity under study with its own individual estimate of error, it is clearly of unquestionable validity.

The commonest cause of unequal variability, however, is dependence of variance upon the mean and the commonest cause of that phenomenon is an unsuitable scale of measurement. The remedy therefore is to "transform", i.e. to use some other scale that will serve better the purpose of analysis, and that often means something quite simple like taking the logarithm of a size to turn it into a growth rate or a square root to change an area into a distance. The intention is to study what is actually varying and not some function of it. Thus, few treatments alter sizes but rather the growth rates that lead to differences in size. The aim is to isolate the biological phenomenon and to measure what is fundamental, not what is derivative.

If that were all, the whole question of transformations could be dismissed as yet another example of the way biological and statistical thinking interact and often illuminate each other. Unfortunately instances arise where the quantity that needs to be measured is determined by practical considerations and information about it is required whether the statistician likes the variate or not. The witticism that farmers are interested in potatoes and not the square roots of potatoes has its point and many have objected to the use of transformations on such grounds. In fact, both sets of needs have to be borne in mind. If the investigator is studying growth, whether measured by trunk circumference or length of shoots, the conclusions must relate to those variates. On the other hand, it is no good going through the motions of analysing data if the conclusions are going to be meaningless. There is, however, a simple resolution of the conflict. So long as the transformation is, to use the mathematician's word, "monotonic", i.e. a change in the untransformed variate implies a change in the same direction in the transformed, there is no irreconcilable difference. For example, an increase in growth necessarily leads to an increase in the logarithm of growth, so an assertion that the one has been significantly increased must mean that the other has been significantly increased also. In fact, all that is required is first to transform the data to form a variate that is statistically acceptable, to work out the analysis of variance and to interpret it, but to effect a "back-transformation" of the treatment means to restore the original scale of measurement before presenting the results. That is to say, if the analysed figures are the logarithms of the original data, then the treatment means will be back-transformed to anti-logarithms for the ordinary reader and similarly treatment means of square roots will be squared before publication.

This simple procedure nevertheless causes difficulty in some people's minds on account of the drift in means brought about by the process of transformation followed by back-transformation. Suppose, for example, that Treatment A gave values of 2, 3 and 4 whereas Treatment B gave 4, 6 and 8, the means being respectively 3 and 6. After logarithmic transformation the values are 0·301, 0·477 and 0·602 for A and 0·602, 0·778 and 0·903 for B with means of 0·460 and 0·761 respectively or 2·88 and 5·77 after back-transformation. The reason is this: the need for transformation arises because the larger values of data have a larger standard error, i.e. are less well determined. When that has been allowed for in the taking of means, i.e. less weight is given to them, the mean value quite correctly drifts towards that part of the range in which values are better determined. That will always happen whatever transformation is used. However, since the drift is fairly constant and depends upon the transformation itself, contrasts are unaffected and no harm has been done, a point of view somewhat difficult to explain to an editor who cannot understand why two sets of means, one down the right-hand margin of a table and the other along the bottom, do not agree. If, further, he is the sort who requires standard errors or significant differences for all means and none appear because they can be applied only to means of the transformed variate, which are not given either, some diplomacy may be needed though everything is in fact correct and reasonable. Short of a

campaign to improve the statistical education of biological editors, there appears to be no solution. A compromise is often found by presenting the means of the original data, uncomplicated by transformation, and the significances, etc., derived from analysing the transformed variate.

Many transformations have been suggested by many people, as is shown by a recent bibliography (Hoyle, 1973), but even a careful survey of the statistical literature cannot be complete. So many data are presented in scientific literature in an unconventional form to meet special circumstances that a comprehensive list is out of the question. Here attention will be directed to transformations that are intended to facilitate statistical studies. The point has already been made that the best transformations do more than that, and isolate the underlying biological feature; there can be difficulties when the two aims are in conflict. On the one hand, there are mathematical expedients that stabilize variances or have some other effect, but only at the expense of generating biological nonsense. On the other hand, there are coefficients of this and indices of that, which are supposed to measure biological parameters but are so awkward in use that a statistician has to return to the original data and start again. Also, they build in a lot of the prejudices of the person who devised them. If nature does behave as he thinks, such measurements may be all right but, if not, their use is confusing and uncertain.

One common transformation is to logarithms (Williams, 1937). The need shows itself when the range of values within each treatment is more or less proportional to the mean of that treatment. However, important as it is for the data themselves to show the need, it is also important that the transformation itself shall be reasonable, which logarithms obviously are when the end-product of a growth process is being considered. Better fertilization gives larger plants because it leads to higher growth rates. A logarithmic transformation of final weights or final trunk circumferences will provide a measure of the rate at which growth took place, which is what was altered by the treatments. It is also what was altered by the differences in environment that led to experimental error. Incidentally, one check on the reasonableness of the logarithmic transformation in any particular instance is afforded by asking whether there can be any zero data, because log (0) is minus infinity. If the answer is that zero is impossible, e.g. a plant must have some weight or it would not exist, all is well. If the answer is that zero is logically impossible, but it does occasionally arise because some very small values cannot be measured finely enough, e.g. a plant's weight may be less than half the unit of measurement and will be rounded off to zero, some adjustment of the transformation may be required in a manner explained below. If the answer is that zero is as good a number as any other and could conceivably arise, a logarithmic transformation is clearly wrong in principle, even though there are no zero data to cause trouble.

The next common transformation is the square root (Bartlett, 1936), which reveals itself if the range of values within a treatment is proportional to the square root of the treatment mean. One case of such a transformation being useful arises when an unlikely event is given many chances of happening. Thus, a leaf is to be sampled for a count of insects. The chances that any particular specimen of the species under study will alight on that particular leaf are very small; on the other hand, the species has many specimens. At a theoretical level it is required that all specimens should act independently. In fact, that is most unlikely. Some species avoid their own kind while others are communal, but such tendencies have to be very extreme before the appropriateness of the square root transformation is seriously disturbed. Lesions caused by fungal spores behave much like insects and also have a tendency to form groups, e.g. from a splash of water. For some reason, blossom counts and records like fruit number or shoot number per tree, also often require a square root transformation. Indeed, it is of wide usefulness. It should be recalled that negative numbers have no real square roots, so the transformation is wrong in principle when applied to a variate that could be negative, e.g. number of blossom clusters this year *minus* the number last year, though no data actually are less than zero. (Incidentally, the example serves as a reminder that transformations are usually best applied to the data themselves, not to derivatives of data, e.g. it would probably

be quite all right to analyse the square root of the number of blossom clusters this year *minus* the square root of last year's number.)

Often in biological work situations arise in which the logarithmic transformation is too powerful and the square root not powerful enough. Usually the distribution is of the so-called "negative binomial" kind (Fisher, 1941), which can lead to useful studies, but the methods are not of a sort to be described here.

A further standard transformation is the angular, i.e.

$$p = \sin^2 \theta$$

where p is the proportion of specimens having a certain property (Bliss, 1937, 1938; Fisher, 1954). As a variate p can be very awkward, being constrained to lie between 0 and 1. At the ends of its range, i.e. if all specimens do or do not have the property, the variance is necessarily zero but rises to a maximum when $p = \frac{1}{2}$. Accordingly some transformation is essential if extremes are found. Unfortunately the angular transformation is no more than a mathematical expedient. Although p, the proportion of cuttings that die or the proportion of buds that develop or whatever else, may make good sense, no meaning can be attached to θ, the quantity analysed. Accordingly the transformation should be avoided where possible. Between 15 and 85 per cent (i.e. $0{\cdot}15 \leqslant p \leqslant 0{\cdot}85$), p and θ lead to such similar results that there is little point in using the transformation at all. Outside that range it is sometimes avoided, but only for simplicity. For example, between 0 and 15 per cent ($0 \leqslant p \leqslant 0{\cdot}15$) a square-root transformation has much the same effect as the angular, while between 85 and 100 per cent ($0{\cdot}85 \leqslant p \leqslant 1$) the square root of $(1-p)$ serves similarly. As long as the data be largely within one of these ranges (0–15, 15–85, 85–100 per cent) the angular transformation is not much favoured, but if they extend, for example, from 8 per cent to 60 per cent it would ordinarily be used.

One difficulty is especially evident with factorial experiments and interactions. If untreated cuttings give a 75 per cent mortality, which reduces to 60 per cent if treatment A is applied or to 50 per cent with treatment B, it would be reasonable to expect 40 per cent mortality if both A and B were used, i.e. if A alone effects a reduction to four-fifths of the original value and B a reduction to two-thirds, it is to be expected that together they would effect a reduction of eight-fifteenths. That would imply the use of a logarithmic transformation to obtain means of

	0	A
0	log (0·75)	log (0·60)
B	log (0·50)	log (0·40)

which show no interaction in the ordinary sense of the word. However, such a transformation is appropriate only when the standard error of the data is proportionate to the untransformed means, which is not the case here. Indeed, the largest mean (75 per cent) is the one nearest to a limit and therefore the least variable. On the other hand, the angular transformation will allow for differing variability but makes the definition of zero interaction even more obscure. There is no completely satisfactory resolution of the problem, which may usefully be taken as showing what happens when a transformation is introduced for a purely statistical reason and not to represent a biological concept more fundamental than that measured by the original data.

One difficulty with transformations arises from "discontinuity". All variates, even if they are naturally continuous, like trunk circumference, are measured only to a certain degree of precision, say, to the nearest millimetre. That does not usually matter so long as the possible values are evenly spaced and not too far apart. However, when those same values are transformed, the increments between them are no longer equal and there can be difficulties, especially with extreme values. Thus, the case has already been considered of plant weights so small that some approximate to zero. So extreme a case would surely call for a reconsideration of the

scale of measurement, but that would be of no avail if the variate were inherently discontinuous, e.g. it represented the number of spores to germinate, the values being 0, 1, 2, etc. and a square-root transformation was required. The usual solution is to add a constant to the variate before transformation, the best value in general being $\frac{3}{8}$ (Anscombe, 1948). Thus, the logarithmic transformation becomes $\log(x+\frac{3}{8})$ and the square root $\sqrt{(x+\frac{3}{8})}$, though it has been suggested that $[\sqrt{(x)}+\sqrt{(x+1)}]$ is even better (Freeman and Tukey, 1949). For the angular transformation p is not taken to be k/n, where n is the number of specimens examined and k is the number to have the property, but $(k+\frac{3}{8})/(n+\frac{3}{4})$.

The whole subject of transformations has been gone into in detail by Cochran (1938) and Bartlett (1947), whose papers, though no longer recent, remain valuable. Clark and Leonard (1939) discussed fully the analysis of data in percentage form. The excellent paper by Box and Cox (1964) shows how diverse are the considerations that could lead to transformations. In this Section all the emphasis has been on "homoscedasity", i.e. equality of variance for all observations, but there are other considerations. Thus, in regression analysis it might be desirable to work with a quantity that bore a simple relationship to the other variates, or it might be important that the transformed variate showed no interaction between factors, as in the example considered above. Again, if it is a question of achieving normality of distribution, Moore (1957) suggested amending the square-root transformation to $(x+\frac{1}{4})^{\frac{1}{2}}$.

Even if homoscedasity is the main character to be achieved, it is possible for variability to depend not upon treatments but upon blocks, a point made in Section 51 when discussing the results of Laycock (1955). The effect is that treatment means should really be modified by giving less weight to those values that come from variable blocks, but that is not easily done. If a transformation will not remedy the defect, it is best to accept an ordinary analysis, noting that its sensitivity is reduced.

A transformation usually helps in achieving "additivity", i.e. the property whereby a plot can be expected to yield according to the sum of two effects, one due to blocks and the other to treatments, with no interaction. However, interactions are possible and their effect has been considered above; tests for non-additivity really amount to tests for certain patterns of interaction. One of the most useful is that of Tukey (1946b) who considered the case with randomized blocks in which treatment means draw apart or come together as block fertility changes but retain their order. His work has been extended to Latin squares by Abraham (1960), to balanced incomplete blocks by Ward and Dick (1952) and to incomplete data by Milliken and Graybill (1971) and Preece (1972). Some general comments have been made by Rives (1962).

Various writers have drawn attention to non-normality of distribution of crop weights, a state of affairs that usually calls for transformation. Thus, Butters (1964) commented on the need for a square-root transformation of coffee yields. For cocoa Pound (1932) and Rogers and Night (1953) found skew distributions, whereas Esquivel and Soria (1967) found the opposite. However, Bartley (1970) concluded that a skew distribution in early years became more nearly normal later, a phenomenon often noticed at East Malling with apple and pear crops. Wellensiek (1938) has commented on the skewness of tea crops. Rives (1960) comments that the structure of the error variance can be as rich as a source of information as the treatment means themselves and points out that graded records usually require transformation.

74. The Analysis of Several Years' Results

One feature of statistical work with perennial species is the accumulation of successive seasons' results for each plot and it sometimes needs a little thought how to make the best use of them.

It should first be commented that there is no need to look for difficulties. Usually it is quite enough to add together all the data from a plot to obtain total crop or total growth, whatever it may be, over a period. With size records, similarly, a figure representing the increment over a period may give all the information required. Trunk circumferences in particular

lend themselves to this treatment. They are usually dealt with in logarithmic transformation and the increment in logarithm of trunk girth is, of course, a measure of rate of growth—a figure of some importance and interest. Usually no more complicated technique need be employed.

If the time taken by the experiment is to be divided into periods, it is best for them to be equal in length and for each to cover an even number of seasons. Most perennial species are to some extent biennial in cropping and growth and, consequently, periods containing an odd number of seasons are rarely comparable one with another.

One useful device for analysing together results for successive periods has already been mentioned (Section 25), namely, regarding the periods as a further factor split upon the smallest plot laid out in the field. Thus, if the trial is already in main plots and sub-plots, the individual results for periods would be regarded as the results of sub^2-plots split upon the sub-plots. The difficulty is that variability in a trial is itself varying, and the resulting heterogeneity of the error of the last analysis (see last Section) cannot usually be dealt with by any transformation. Also the years are best regarded as a treatment applied systematically, so the appropriate method of analysis is that given in Section 46.

There is much to be said for the following method, used with excellent effect by Stevens (1949) with coffee. Suppose there are two periods into which the trial has been divided, which may be designated by (i) and (ii). First of all, an analysis can be done on (i)+(ii), which will give a measure of the total yield or growth or disease, as the case may be, over the whole experimental period. Then a separate analysis can be done on (ii)−(i), representing the build-up of the variate with time. Again, if there are three periods, analysis can usefully be done on (i)+(ii)+(iii), (iii)−(i) and (i)−2(ii)+(iii), representing respectively the total over the experiment, the degree of increase during the experiment and the extent to which the middle period gave results above or below the mean of the other two. This last can be important for, if results from the second period are much above the mean, the conclusion is that the quantity, having built up quickly, is later increasing more slowly. Thus, if the figures for crop over three periods are 10, 30, 35, the suggestion is that large further increases in cropping are not to be expected, whereas figures like 10, 12, 35 give hope of larger crops for the future.

A special case of this approach is provided by working out an analysis of variance on the regression coefficient of the variate upon time. Thus, suppose crops over four successive periods for a certain plot are 10, 13, 15, 16. Using the ordinary designation of the periods as 1, 2, 3, 4, this gives a regression coefficient of (see Appendix IV)

$$\frac{(10\times 1)+(13\times 2)+(15\times 3)+(16\times 4)-\frac{1}{4}(10+13+15+16)(1+2+3+4)}{1^2+2^2+3^2+4^2-\frac{1}{4}(1+2+3+4)^2}$$

The denominator may be ignored as being the same for all plots. Further, the numerator can be simplified by renumbering the periods in such a way that their numbers are still evenly spaced but add up to zero. Thus, in the present example, renumbering the periods -3, -1, $+1$, $+3$, gives the expression,

$$(10\times -3)+(13\times -1)+(15\times +1)+(16\times +3)$$

That is just as good as the regression coefficient for showing the relative degree to which plots are giving an increase in crop as time goes on.

In the investigation of Stevens, already referred to, each plot gave a crop record for ten successive years. First, he analysed the total crop,

$$\text{(i)}+\text{(ii)}+\text{(iii)}+\text{(iv)}+\text{(v)}+\text{(vi)}+\text{(vii)}+\text{(viii)}+\text{(ix)}+\text{(x)}$$

Then he analysed the difference in cropping between odd and even years to investigate the degree of biennial bearing,

$$\text{(i)}-\text{(ii)}+\text{(iii)}-\text{(iv)}+\text{(v)}-\text{(vi)}+\text{(vii)}-\text{(viii)}+\text{(ix)}-\text{(x)}$$

Finally, he grouped his years into pairs in order to remove the biennial effect and worked with a regression coefficient, as modified above, to find the extent to which cropping was improving with time, i.e. he analysed

$$-2(\text{i})-2(\text{ii})-(\text{iii})-(\text{iv})+0(\text{v})+0(\text{vi})+(\text{vii})+(\text{viii})+2(\text{ix})+2(\text{x})$$

In that way he studied the important effects associated with time without artificiality and without making any dubious assumptions about the homogeneity of errors.

The method has found extensive use (e.g. Morales, 1951; Pimental-Gomes, 1963) and has been especially recommended (Fraga and Conagin, 1956) for use with old experiments designed before modern ideas were appreciated.

The above example neatly illustrates a point that is not always apparent, namely, that there is a close relationship between split-plots and partitioning. The simplest case arises when there are two varieties, representing perhaps sampling on two occasions or by two people. One possibility is to analyse as a split-plot design with two sub-plots to a main plot. Another is to analyse first the sum of the two values and then their difference. It will appear that the analysis of the sums is the same as the main-plot analysis except that all sums of squared deviations are doubled, while the analysis of the differences corresponds similarly to the sub-plot analysis. If each variate had been divided by the square root of two, two being the sum of squared coefficients in the functions (1, 1) and (1, −1) (see Section 35), the correspondence would have been exact. Similarly, if there had been three samples or people or whatever, analysis of $(1, 1, 1)/\sqrt{3}$, $(1, -1, 0)/\sqrt{2}$ and $(1, 1, -2)/\sqrt{6}$ would have provided a split-plot analysis, the first analysis being equivalent to the main-plot analysis, while summing of corresponding lines in the other two would have given the sub-plot analysis. The point is worth noting, even if only because it provides a computational device; it shows that an experimenter can always produce a split-plot analysis if he can analyse data from the main-plot design. Accordingly, computer programs for split-plot designs though useful are not essential. The advantage in the present context, however, is not merely computational. For one thing, it is not necessary to analyse all contrasts between years but only those that look interesting; for another, if trouble is feared from differences of variability between the years, there is no need to add the analyses of contrasts together, thus pooling the errors.

Quite a different approach is to regard successive years' crops as separate variates and to use multivariate methods (see Appendix II) as advocated by Steel (1955) with Finney (1956) dissenting. The approach looks promising, but it does not appear as yet to have had any noticeable successes. It has, however, been used (Pearce and Holland, 1960).

One special problem associated with perennial species is that of biennial or other cyclic bearing. In one year the tree crops heavily; in the next its activity lies chiefly in growth; in a third year it returns to cropping and so on. If the alternation were completely regular the method described above as used by Stevens (1949) would suffice, but changes of phase complicate everything. Also, as Haines (1958) and Chapas (1961) have noted for oil palms, the period can be longer than two years.

Hoblyn, Grubb, Painter and Wates (1936) suggested that biennial phenomena be considered under two heads. First, the fact of cyclic bearing should be established by the evaluation of B, which is derived thus:

Successive crops are	10	15	5	15	25
The changes are		+	−	+	+
Alternations?			Yes	Yes	No

Hence there are two alternations out of a possible three and $B = \frac{2}{3}$. The measure is, however, rather insensitive. In economics it is known as the "turning point test" and Kendall (1948) has pointed out, if there are N years, given a random series the expected number of alternations is $\frac{2}{3}(N-2)$, so in the above example the value of B is only what would be expected if crops

occurred at random with no pattern, biennial or otherwise. The variance of the number of alternations is $\frac{1}{90}(16N-29)$, so, if there were nine years with a perfect biennial effect, the number of alternations would be 7 against an expected number of $4\frac{2}{3}$, the standard error being $\sqrt{(\frac{23}{18})}$, i.e. the biennial effect would be about significant at the level $P = 0{\cdot}05$. Of course, it is possible to work out B for each plot of an experiment and to use the analysis of variance but the insensitivity will remain. If cropping potential is not constant but tends to rise, the number of positive changes may be expected to increase, which makes alternations less likely to arise.

The same writers (Hoblyn *et al.*) proposed a second quantity, I, to measure the intensity of a biennial effect, supposing one to have been established. It is derived from the differences in successive crops in relation to their sums. Thus, in the example given above

$$I = \tfrac{1}{4}(\tfrac{5}{25}+\tfrac{10}{20}+\tfrac{10}{20}+\tfrac{10}{40}) = 0{\cdot}29$$

Both B and I were considered by Abeywardena (1962a) along with other measurements with special reference to coconuts.

A later study was made (Pearce and Doberšek-Urbanc, 1967) using a computer to simulate actual situations by building up a pattern (i.e. with or without a trend, with or without additional biennial or triennial effects) and adding a random variable to represent error. For each of the sequences of successive "crops" thus generated, some alternative measures were tried. The authors begin their discussion with the following words:

> "The main difficulty in evolving a measure of biennial phenomena is one of definition. If experimenters were concerned only with a strict alternation of high and low crops according to some mathematical law, the task should be easy. In practice, however, an occasional change of phase is held to detract little from the intensity of bienniality. Also, a tendency to periodicity with longer intervals is regarded as a diluted biennial effect; at the extreme, non-periodic variation, i.e. randomness, can still be regarded as an aspect of the same phenomenon. Where the definition is so loose, no precise measure is possible. The sole inconstancy of cropping or growth that does not qualify as bienniality is a steady increase or decrease, though it is hard to think of any measure of bienniality that will be unaffected by it."

They concluded that I is reasonably satisfactory even when used without B. They also found advantages in a quantity, K, derived by squaring the components of I, i.e. in the example

$$K^2 = \tfrac{1}{4}[(\tfrac{5}{25})^2+(\tfrac{10}{20})^2+(\tfrac{10}{20})^2+(\tfrac{10}{40})^2] = (0{\cdot}39)^2$$

but did not consider it so much better than I as to justify a change of established practice. Whether I or K is used, values should be worked out for all plots, and examined by the analysis of variance.

Quite a different approach was suggested by Dancer (1963), who proposed judging the fruitfulness of a plot by the regression coefficient of its cumulative crops on age.

75. Experiments at Several Sites

Here, as in the last Section, there is no need to look for difficulties. However, clarity about presuppositions and objectives is essential and many co-operative experiments and series of experiments fail because inconsistent decisions are made. The first point needing clarification is the likelihood of an interaction between sites and treatments. If not agreed right from the start that results may be quite different on the clay soils of the south-west and the light loams of the north-east, an experiment is required in each area, sufficiently large to reach a definite conclusion. If, on the other hand, it is assumed that results will be much the same wherever the trial is placed, it could be sufficient to put down one block here and another block there and to analyse them altogether as a single experiment, the replication being that of the exercise as a whole. What is wrong is to include two or three replicates at each site "in case there is an

interaction", unless the replicates are numerous enough to give an answer for each site separately if the need should arise. Far too many experiments fall between two stools; they test for an interaction and invalidate themselves if one is found.

However, the analysis of data from several sites often encounters another difficulty, namely, that of differing error variances. If one part is planted in good conditions and another in bad, initially at least the second lot of trees will be more variable, though they may even up later, but if one site is on a slope and the other is level, differences due to water relations or exposure might build up rather than disappear. The problem is this: if a difference of a is found between two treatment means at one site with a variance σ^2, while at another site a difference of b is found with variance $2\sigma^2$, what estimate should be made of the real difference? The difficulty has to be considered in relation to the former one. If an interaction is expected, the answer is obviously a at the first site and places like it, but b at the second. If, however, it is an article of faith that there is no real difference between the two sites, the answer is $\frac{1}{3}(2a+b)$ for the two sites taken together, since more weight should obviously be given to the site where a better determination was made.

The trouble is that a lot depends upon the extent to which anyone believes that the sites are really equivalent. If they are as similar as everyone says, how did it come about that they gave different variances? If someone replies that planting at the second site took place during a drought, a more reasonable reaction might be to weight a and b by the frequencies with which trees are planted in good conditions and in bad, values that are however probably unknown. There is something artificial about weighting on statistical grounds alone and ascribing special importance to one site without asking what the weights really represent.

Of course, if the variance was double at the second site because there was only half the number of replicates, everyone will agree to the other being given double weight. The present plea is for circumspection. The formulae given in textbooks are all right where correctly applied. They can be all wrong if applied without thought.

Also, there are contexts in which an average over an area is not required. On the other hand, where there are bulk purchases of plants, fertilizer or spray materials by a co-operative or the government, general recommendations have to be made, but site-treatment interactions still need to be considered to avoid some unfortunate growers, who have to farm in specialized conditions, being harmed. The conclusion for them might be to withdraw from the co-operative or to make their own purchasing arrangements or even to try a different crop, but they must have warning that the group decision will not favour them.

Much has been written on the subject and the statistical aspects have been well explored (Cochran, 1937; Roessler and Leach, 1944; Cochran and Cox, 1950; Cochran, 1954). Study of these references could be very worth while for anyone faced with a real problem. As long as he sees his own problem clearly all will be well, but this is not a field in which mathematical virtuosity pays dividends. If doubt arises whether variation is different at different sites, the non-technical paper of Roessler (1943) may be found helpful.

CHAPTER 8

SOME MISHAPS AND REMEDIES

". . . of most disastrous chances,
Of moving accidents by flood and field."
WILLIAM SHAKESPEARE, *Othello*, I.iii.136–137.

80. Errors and Blunders

A field experiment requires much thought and supervision if all is to go well, so the practical aspects of organization are of the highest importance. Planting can easily go wrong, the application of treatments calls for great care, while the maintenance of good standards of harvesting and crop recording is not easy, especially as operations sometimes take place in a rush and often with uninstructed casual labour to complicate matters. The problems have to be foreseen, so there must be close and willing co-operation between farm foremen, who cannot be expected to see all the scientific implications of practical decisions, and the experimenter, who must be prepared to get his hands and his boots dirty as he keeps an eye on what goes on. Each should recognize and value the skill of the other. A postgraduate degree in botany is not intended as a qualification in the practicalities of field experimentation and its owner should understand that he still has much to learn. The worst results are obtained when the experimenter recognizes neither the role he should play nor his dependence upon others and withdraws from the scene declaring that everything is in the hands of experts who know their job and can be left to do it. That is as bad as the parallel faults of leaving the data to be analysed by computer scientists and asking a statistician to design an experiment without telling him what it is required to do. Experts can help only when they know what is expected of them. When things go wrong the blame must attach to the man at the top, usually to some highly qualified scientist who had his head in the clouds when he should have been conferring with his helpers about practical matters.

Nevertheless, whatever precautions are taken, things will go wrong occasionally. Tukey (1963) wisely comments that there are errors and there are blunders; the first must be lived with and the second avoided. He adds that the distinction "is partly a matter of size of deviation, but more a matter of difference in the character or assignability of causes". In practice there will always be the occasional blunder, while serious errors will have much the same effect. The learned theoretical statistician, who wrote in his book that he would not consider the consequence of incomplete data because at well-conducted research institutes mistakes do not occur, may have thought that he was avoiding a difficult subject and doing so in a manner that would ingratiate himself with his reader. In fact, he made an ass of himself.

One cardinal principle in dealing with defective experiments or defective data is to ask whether the trouble derives from an irrelevant external factor, as when a truck plunges off the road and knocks down some of the trees, or from a relevant internal one, as when a variety in a trial of new seedlings shows that it has no resistance to a disease that is endemic in the neighbourhood. Put like that the distinction looks obvious; more usually it is made only with difficulty. In general, it is safer to blame the treatment for damage. Thus, in a trial of tree-forming methods, branch breakage could arise from defects carried over from the nursery, but it would be wiser to let the treatment carry the blame rather than make some adjustment to restore the position.

Sometimes the trouble is of such rare occurrence that a treatment is not thought worse of because it makes the plants susceptible to that kind of damage. Thus, in areas where spring frosts are almost unknown, blossom injury to an early flowering variety might be regarded as a misfortune, irrelevant to any assessment of the variety. Again, quarrying operations may cause dust injury, but it would not be fair to penalize a treatment that proved to be

especially susceptible. That is not to say that trials can be satisfactorily carried out in such unusual conditions.

Where allowance for injury is to be made, there are several ways of doing it. One is to leave whole blocks, rows or columns out of all analyses of variance, supposing the design is robust enough to allow that (see Section 21). Certainly, the surest method of eliminating the effect of some interfering factor is to ignore data from the blocks where it operates. For that reason the recommendation has already been made (Section 54) to plant more replicates than will be needed if all the trees remain available for study. Usually, however, that method of allowing for injury is not convenient because the infected area spreads over too many blocks. The remedy then may be to measure the injury and adjust the results by the method of covariance (see Appendix IV) or to regard certain plots as missing or defective.

81. Subsequent Classification and Pseudo-variates

Somewhat similar problems arise when it is found too late that the trees differ in some way not considered while the trial was being designed. Thus, a set of rootstocks supposedly from the same clone might be found to consist of a mixture of two identifiable clones. Sometimes, too, a guard row has to be removed on account of the needs of an adjacent experiment, thus classifying the remaining trees into "inside" and "outside". The best way of dealing with such mishaps is to associate each tree with a *pseudo-variate* (e.g. Quenouille, 1948), which is 0 if the tree belongs to one class and 1 if it belongs to the other, to add up the tree values to give plot values and to adjust by the analysis of covariance.

Thus, suppose there are four rows each of five trees in an experiment and the north and east headlands become unguarded. Since the two headland effects may well be different, each must have its own pseudo-variate, thus,

1	1	1	1	1	1
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0

0	0	0	0	1
0	0	0	0	1
0	0	0	0	1
0	0	0	0	1

and a double covariance will be needed.

In working out adjusted means, it is usually more sensible to adjust on to a standard value of zero for the independent variate rather than on to its general mean, i.e. to adjust all treatment means to the value of x representing the rootstock that should have been planted or the circumstance of remaining guarded rather than to the mean of what actually occurred. This makes a constant difference to all the adjusted treatment means but does not alter the computations set out in Appendix IV in any other way.

Pseudo-variates can also be used at times to take account of soil variation. Thus, in one raspberry variety trial designed at East Malling difficulty was experienced from a streak of gravel that lay across the area. Plots on good land were assigned the value 0, plots wholly on the streak were assigned the value 1 and a few that lay on the boundary were given values like $\frac{1}{4}$, $\frac{1}{2}$, according to the proportion on gravel. Adjustment by covariance led to a marked improvement in precision.

82. Missing Trees and Plants

If a tree dies or has to be removed on account of disease, it is advisable to put another in its place, the new one being known as a "replant", a "supply" or a "recruit". The consequences when a *whole* plot is thus lost will be discussed in the next Section, this one dealing with incomplete plots.

If the losses are probably due to the effects of some of the treatments, it would be wrong to make any adjustment. That is to say, trees should be replanted as occasion arises and

analyses should be carried out on plot totals without adjustment of any kind for the number of gaps or replants, the crop, etc. of the replants being counted in with the other trees.

If the effect of missing plants is to be allowed for, there are two ways of attempting it, but first it should be mentioned that the effect of gaps on adjacent plants depends very much upon the species. Thus, for strawberries the effect, if any, is small (Taylor, 1951b), but for hops it can be appreciable (Amos and Hoblyn, 1928). It has been suggested that it is important for pecan trees also (Sharpe and Blackmon, 1950).

The best way of allowing for the loss of plants is to use the analysis of covariance, adjustment being necessary both for the number of gaps or replants and for the effect of competition. Two alternative methods suggest themselves:

(1) To work with *plot totals*, ignoring replants, and to adjust upon the number of gaps or replants as the case may be. This is open to some criticism on the ground that the loss of a plant from a plot that is doing well is more serious than from one that is doing poorly, thus vitiating the assumption of the analysis of covariance that the regression coefficient is not affected by the treatments.

(2) To work with *the mean per surviving plant* for each plot, again ignoring replants and again adjusting by the number of gaps or replants. The advantage of this method lies in the number of survivors having already been taken into account, so the experimenter is free to concentrate on devising an independent variate that shall express the competition between plants. Thus, he may weight gaps by the time since their occurrence, assigning a value 5 to one that appeared five years ago and ignoring one that occurred yesterday. Again, given three gaps in a row he may ignore the middle one as having no effect on the competition of the survivors. Admittedly, he must not become too ambitious and lose himself in a welter of trivialities and he must remain objective, not seeking reasons why his favourite treatment was really best despite all appearances, but he is much freer to devise a useful independent variate. The criticism of the first method applies here also and is difficult to counter, though it should be remembered that the differences most seriously affected are the large ones, which are obvious anyway.

Generally it is advisable to replant as soon as possible in order to restore root competition for the adjacent trees; but, where most losses occur while the trees are young, it is sometimes better to wait until no more are expected and then to replant all the gaps together. The replants are now comparable with one another, so their crop, etc. may be counted in with the plot totals, the difference between them and the original trees being dealt with by a pseudovariate.

In some crops, e.g. cocoa, the number of trees to a plot of standard area can be very variable in commercial holdings. In such instances it is advisable to work with plot totals and to adjust by stand, supposing the latter to have been determined before the treatments were applied. The reason for preferring the first method to the second is that, with no standardized planting distance, attention has to be fixed on results per unit area. This is not really within the province of this Section though it is a closely related subject.

The difficulty with all these methods is their tendency to give heterogeneous errors (see Section 73) that no transformation can deal with. If one plot has 10 trees surviving out of 12 and another only 2, it is absurd to suppose that the accuracy of the results for one plot equals that of the other. If too many plants are lost from a plot, it is best regarded as wholly missing.

83. Missing Plots

Although the loss of complete plots from an experiment is not a disaster and, with some species, is quite usual, it should be emphasized that *nothing can replace the lost information.* It is a bad mistake to suppose that a few simple modifications in the computing will set all to rights. For one thing, the modifications are not always simple; for another, something must always be ost from the precision of treatment comparisons.

One common method of dealing with incomplete data is to fit "missing plot values", contrived so that the plots concerned contribute a zero residual to the error sum of squared deviations. Alternatively, with block designs at least, it is usually possible to analyse the data as they stand, using one of the modern general methods (e.g. that in Appendix III). Yet a third possibility is to assign an arbitrary value to any missing plot and to adjust by covariance upon pseudo-variates (Bartlett, 1937; Nair, 1940; Smith, 1950; Coons, 1957). If one plot is missing, one pseudo-variate will be needed, with a value of 1 for the plot that is missing and of 0 for all others. If several are missing, one such pseudo-variate will be needed for each. These different methods will all lead to rather different results. The situation may be a little perplexing to the uninitiated, but it is quite simple really, as will appear below. In the case of a row-and-column design, a general method of analysis will be complicated on account of the loss of orthogonality between rows and columns, but the other two methods remain available.

The first perplexing point, if all three methods are used and the results compared, is that the treatment means given by missing plot values and the general method are not the same, though differences are preserved. This effect was first noted by Gilbert and Grossmann (1961) and arises from the different areas to which mean values are being referred. With the missing plots represented by calculated values a mean indicates what would have happened if the whole area had been given that treatment; with the missing plots excluded the corresponding mean indicates what would have happened if the surviving plots had received it. Which is better depends upon the experimenter's assessment of the reason for some data being missing, but since experiments estimate the differences between means rather than the means themselves the point is rather academic. Anyway, the discrepancy is likely to be small and the choice of procedure will in practice depend upon other considerations. The third method, that of covariance, will give means that depend upon the values to which the pseudo-variates are adjusted. If all are set equally to zero, i.e. the value for existent data, the results will be the same as if missing plot values had been used.

The next perplexity concerns the analysis of variance. Analysing what remains or analysing the whole and adjusting by covariance will give the same treatment sum of squared derivations, but the use of missing plot values will make it rather larger. Here the analysis using missing plots is wrong, because it has failed to allow for the fact that different missing plot values would have been obtained if it had been decided beforehand that treatments were having no effect and could be merged with error, a possibility implied by the use of a significance test. If no F-test is proposed, there is no need to adjust the treatment line of the analysis. If the F-value is very large or is too low for significance, there is again no need to adjust, but if conclusions would be different if the treatment sum of squared deviations were somewhat smaller an adjustment is called for. The data have to be analysed again omitting the treatments, thus reducing a two-way classification to one-way, a three-way classification to two-way and so on. As a result a larger error sum of squared deviations will be found and its excess over that for the full design will correctly be that for treatments.

A third perplexity concerns the covariance matrix of the treatment means, which can look very queer indeed when data are analysed as they stand without reference to the original design. However, that is only another aspect of the general shift of the means. Thus, those that were independent before have now suffered a similar shift and have become correlated and so on. Once the variances of contrasts are retrieved they are found to be the same as for other methods. Further, where two treatments have escaped loss of data, the variance of the difference between their means is the same as for the complete design originally planned (Pearce and Jeffers, 1971). Again, the difficulty is more apparent than real.

In all these methods it is necessary to reduce the number of degrees of freedom for error by the number of missing plots and the same, naturally, applies to the total number of degrees of freedom. Where the data are analysed as they stand, the adjustment is made automatically and it is made also using the covariance method, because each independent variate absorbs a

degree of freedom as a matter of course. It is necessary to remember the reduction only when missing values are to be fitted.

Where missing plot values are required, there are two main approaches. One is by formulae That was the original method of Allan and Wishart (1931), though the reader should be warned that there is a slip in their expression for a Latin square. The books of Cochran and Cox (1950) and Federer (1955) are notable for the thoroughness in presenting such formulae, many of them derived from the work of Cornish (1940a, b, c, 1941). Also, Baten (1943) has given expressions for the case when two or three plots are missing from a design in randomized blocks.

In general, however, it is not difficult to find the formula if it is feasible to write down a residual in explicitly algebraic terms. It then requires only to be equated to zero.

The other approach, due initially to Yates (1933), is to find the missing plot values by iteration, a process explained in Appendix III. First some approximations to the missing values are proposed, whether by guess or in some more formal manner is not for the moment important. They are included with the data, which are then analysed as far as the determination of the residuals for the missing plots. If, by great good fortune, they are all zero, the process need go no further, but if not, the missing plot values are amended in the light of the residuals and the calculations are repeated until zero residuals are eventually obtained.

It will here be convenient to illustrate the approaches with an example. Suppose that two plots are missing from an experiment in randomized blocks, thus:

	Treatment				
	A	B	C	D	E
Block I	14	M_1	19	18	19
II	15	14	19	16	M_2
III	14	12	16	15	18
IV	16	17	20	16	21

First, to use a formula, the residual in a randomized block design equals

plot value − the expectation of the plot value

= plot value − [the general mean + an adjustment for blocks + an adjustment for treatments]

= plot value − [(general mean) + (block mean − general mean) + (treatment mean − general mean)]

= plot value − block mean − treatment mean + general mean.

Therefore, taking the two missing plots separately,

$$0 = M_1 - (70 + M_1)/5 - (43 + M_1)/4 + (299 + M_1 + M_2)/20$$

$$0 = M_2 - (64 + M_2)/5 - (58 + M_2)/4 + (299 + M_1 + M_2)/20$$

that is,

$$12M_1 + M_2 = 196$$

$$M_1 + 12M_2 = 247$$

so

$$M_1 = 14{\cdot}7, \quad M_2 = 19{\cdot}4$$

It is a sound computing practice to calculate missing values to one place of decimals more than the data.

Next, to consider iteration, let the two missing values be assigned guessed values, say, the means of existent data in their respective treatments, i.e. $M_1 = 17{\cdot}5$, $M_2 = 16{\cdot}0$, giving residuals of $1{\cdot}5$ and $-1{\cdot}9$ respectively, which show that the guesses were not altogether successful.

The simplest procedure is that of Healy and Westmacott (1956), who suggested merely subtracting the residuals from the guessed values and trying again, i.e. with 16·0 and 17·9, which gave residuals of 0·7 and −2·0, leading to 15·3 and 19·9 as the next pair of proposed values and so on. It was later suggested (Pearce, 1965a; Pearce and Jeffers, 1971; Preece, 1971) that the process would be speeded if the residuals were modified before subtraction by multiplying each by the number of plots for the complete design and dividing the product by the error degrees of freedom, thus changing 1·5 into $1{\cdot}5\times\frac{20}{12}=2{\cdot}5$ and −1·9 into −3·2. The values initially guessed, 17·5 and 16·0, would then become 15·0 and 19·2 for the second cycle and so on. When zero residuals are obtained the missing plot values have been found.

The working in the above example has been fairly simple because a design in randomized blocks provides an easy way of calculating the residuals, but that is not always the case. It is therefore useful to know the device of Rubin (1972). He proposed, in effect, using the computer program for the original design to analyse the pseudo-variates that would be used in the covariance method of adjusting for losses. Thus, in the above example he would have analysed:

$$\begin{matrix}0&1&0&0&0\\0&0&0&0&0\\0&0&0&0&0\\0&0&0&0&0\end{matrix}\quad\text{and}\quad\begin{matrix}0&0&0&0&0\\0&0&0&0&1\\0&0&0&0&0\\0&0&0&0&0\end{matrix}$$

For the first pseudo-variate the residuals corresponding to M_1 and M_2 are $\frac{12}{20}$ and $\frac{1}{20}$ respectively; for the second pseudovariate they are $\frac{1}{20}$ and $\frac{12}{20}$ respectively, which may be written as a matrix (see Appendix II) thus:

$$\frac{1}{20}\begin{pmatrix}12&1\\1&12\end{pmatrix}=\mathbf{A}$$

Following Rubin's method **A** requires to be "inverted", a procedure that will not be described here but many computer programs exist to carry it out. (As will be explained later, the operation can be avoided.) However, those familiar with the calculations will readily conclude that the inverted matrix is

$$\frac{20}{143}\begin{pmatrix}12&-1\\-1&12\end{pmatrix}=\mathbf{A}^{-1}$$

Arbitrary values are now assigned to the missing plots and residuals are evaluated (Rubin himself puts all missing values equal to zero, which simplifies the algebra, but computationally there are advantages in the arbitrary values being near the true ones). In this instance it is already known that the arbitrary values of 17·5 and 16·0 lead to residuals of 1·5 and −1·9, so that provides a convenient starting point. The true missing-plot values can be found by making the following adjustments, the coefficients being derived from $\mathbf{A}^{-1}$.

$$\tfrac{20}{143}[(12\times1{\cdot}5)+(-1\times-1{\cdot}9)]=2{\cdot}8$$

$$\tfrac{20}{143}[(-1\times1{\cdot}5)+(12\times-1{\cdot}9)]=-3{\cdot}4$$

Subtracting these adjustments from the arbitrarily chosen values gives the missing-plot values of 14·7 and 19·4. Alternatively, if there are reasons for avoiding the matrix inversion, e.g. the experimenter does not know how to invert matrices, a further iterative method is available (Pearce and Jeffers, 1971). The diagonal elements of **A** are worked out, i.e. $\frac{12}{20}$ in each instance, and are used to divide the respective residuals before they are subtracted. In this instance the method is the same as one already described because $\frac{20}{12}$ is the ratio of the number of plots to the number of error degrees of freedom; in fact, the approximate method given above does happen to be exact in many straightforward situations. With complicated designs, however, substantial gains can arise from using the diagonal elements of **A**.

The special problems raised by missing plots in regression and covariance analysis are not great. With such techniques much depends upon having comparable data from all variates, so if a value is missing from one variate the corresponding values must be discarded from the others, even though nothing is wrong with them, because no use can be made of them. Hence, when covariance adjustments have to be made in an analysis of variance, a plot having data defective in one variate must be regarded as completely missing. If missing values are to be found, they can be fitted for each variate separately as if it alone were being analysed. If the missing plots are being dealt with by covariance, the same pseudo-variate will serve for all, e.g. if it is intended to adjust y by x with two plots defective, it will suffice to adjust y by x and two pseudo-variates.

However, an important qualification needs to be noted. It sometimes happens that values are missing from the dependent variate but not from the independent. In that case, as has been explained, for purposes of determining regression coefficients, etc. defective plots must be regarded as wholly missing, but when it comes to deciding the values of the independent variate to which adjustment is to be made, the experimenter could reasonably assert that he knows the mean value correctly and he wishes everything to be adjusted to it. If that is what he wants, there is no need to object.

84. Some Consequences of Fitted Values

It was mentioned in the last Section that values fitted to minimize the error lead to an overestimation of the sum of squared deviations for treatments. Strictly two sets of missing values should be obtained. First, values should be found to minimize the error; then treatments should be ignored, leaving one classification instead of two or two instead of three or whatever the case may be, and different values fitted. Thus, in the example above it was found that missing-plot values of 14·7 and 19·4 minimized the error. In fact, they led to this analysis of variance

Source	*d.f.*	*s.s.*
Blocks	3	23·15
Treatments	4	77·73
Error	12 − 2 = 10	10·79
Total	19 − 2 = 17	111·67

However, once it is decided to test whether treatments have an effect or not, it becomes necessary to ignore them and see what happens. The result is a one-way classification by blocks alone, the missing-plot values, however found, coming to the means of the existent data in the blocks made defective, i.e. 70/4 = 17·5 and 64/4 = 16·0. With these values the analysis becomes

Source	*d.f.*	*s.s.*
Blocks		28·44
Error including treatments	16 − 2 = 14	73·00
Total	19 − 2 = 17	101·44

The true sum of squares for treatments is 73·00 − 10·79 = 62·21 with (14 − 10) degrees of freedom, not 77·73 as given by the upper analysis. The correct value is given as a matter of course by the covariance method and an analysis of the data as they stand, because the double minimization procedure is implicit in those two approaches.

Another place where care is needed with fitted missing-plot values is where variances come to be considered. Again, both the other methods give correct results but there is a tendency

with fitted values to proceed as if nothing had happened, which is of course wrong. If a mean is based on defective data and a variance is required for a difference that involves it, the best procedure possibly is to use the method of Appendix III to find the covariance matrix of adjusted treatment means and so to derive the variance required. It is true that there is a drift on all means so the method could not be used to find the variance of a mean in isolation, but a change of origin does not affect differences. For randomized blocks Baten (1952) has given formulae for the variances of means affected by missing plots. The matter has been considered also by Yates (1933), Taylor (1948) and Das and Kulkarni (1958).

After these comments on its weaknesses there might seem to be a case for never fitting missing-plot values but for always using some other method. In fact, missing values are ordinarily worked out, partly on account of simplicity but partly because the procedure minimizes disturbance to means for which data are complete.

85. Mixed-up Plots

It sometimes happens that the yields of two plots become mixed and it is not possible to say how much came from each, though there is no doubt about the total. It is sufficient (Nair, 1940) to proceed as if the crop came from one plot and nothing from the other and then to adjust by covariance upon a pseudo-variate that has values of 1 and -1 for the two muddled plots and 0 for the rest. A similar difficulty arises when two samples for analysis become confused so that it is not possible to be certain which belongs to which of two plots. If both are analysed and twice the mean is ascribed to one and zero to the other, a similar adjustment can be used.

If several plots are confused, more pseudo-variates are required. Thus, if three plots are involved in the tangle, a first pseudo-variate, which ascribes values of 1, -1 and 0 to the three plots and zero to the rest, requires to be complemented by another that ascribes values of 1, 1 and -2 to the mixed-up plots and zero to the rest. A double covariance adjustment on the two pseudo-variates will provide a valid analysis. If four plots were mixed up, three pseudo-variates would be required, ascribing respectively $(1, -1, 0, 0)$, $(1, 1, -2, 0)$ and $(1, 1, 1, -3)$ to the four plots involved, all others being ascribed zero in each pseudo-variate, and so on for more mixed-up plots. An alternative but equivalent set of pseudo-variates has been proposed by Das and Bhargava (1958).

Such groups of pseudo-variates can be used in combination, if need arises, along with those for missing plots. Thus, if six plots are defective, A being missing, B and C mixed up and D, E, F mixed up, four pseudo-variates would be needed. Each would ascribe zero to plots not involved and they would ascribe $(1, 0, 0, 0, 0, 0)$, $(0, 1, -1, 0, 0, 0)$, $(0, 0, 0, 1, -1, 0)$ and $(0, 0, 0, 1, 1, -2)$ to plots A–F respectively. A quadruple covariance adjustment upon all of them would provide a valid analysis. The advantage of having a computer program to calculate covariance adjustments for a good number of independent variates is very apparent.

Other ways of dealing with mixed-up plots have been suggested (Bose, 1938; Bose and Mahalanobis, 1938) and latterly attention has turned to iterative methods (Preece and Gower, (1974).

86. Errors in the Application of Treatments

The Sections above have dealt with faults in the data, but it could well happen that something has gone wrong with the design. For example, in a spraying trial with two factors, early (E) and late (L) applications, it could be that a late application is given to a plot that should have had only an early one, turning it from treatment E to EL. The muddle could now take one of two turns. The operator may not apply the spray to the plot that should have received it, thus turning it from treatment L to O, or he may attempt to keep everything according to plan and so keep it as treatment L. Either way the statistician may have a problem, but with the general methods of analysis now available (see Appendix III) he should be able to work out the analysis of variance correctly. His difficulties will come, however,

when he tries to partition the treatment sum of squared deviations into main effects and interaction. The situation is not past redemption but mostly a solution will have to be found by working out the adjusted means for the four treatments, O, E, L and EL, and their covariance matrix and recalling that, in the notation of Section 35, the main effect of E is estimated by (*EL−*L+*E−*O), that of L by (*EL+*L−*E−*O) and the interaction by (*EL−*L−*E+*O). It is unlikely that all three will be estimated independently, but that is the best that can be done.

The difficulty, however, was in the partitioning of the treatment sum of squares. Where there is no need to partition, everything is much simpler. Thus if treatments A and B are interchanged so that A occurs twice in Block I but not in Block II while B does not occur in Block I but does so twice in Block II, the result may be less than ideal but at least a completely correct analysis can be obtained using the algorithm in Appendix III. Sometimes, of course, an interchange takes place entirely within a block and then everything is quite simple.

The most robust case is that in which a treatment occupies the same proportion of the plots in each block. It is then said to be "orthogonal to blocks" (Section 30). Where a number of treatments possess the property, their means and the differences between them are all estimated with full efficiency, i.e. the expressions given in Section II apply. Clearly they will retain the property whatever muddles are made in the allocation of other treatments.

At the other extreme, if a design is disconnected, i.e. if a set of treatments occurs only in certain blocks, the other treatments occurring only in the other blocks, there is no information about the contrast between the two parts. (An example is afforded by confounded designs. Here one side of a contrast is in some blocks and the other side in the rest, thus sacrificing the information about an interaction. See Section 35.) Hence, mishaps that lead to disconnection can be disastrous.

Another important property that may be preserved despite misadventure is that of "proportionality", which is explained in Section 30. The following design is so arbitrary and chaotic that it would appear to have no pleasing features:

Block I	A	B	B	C	D	
II	A	A	B	B	B	B
III	C	C	D	E		

Nevertheless, in each block B always appears twice as often as A, so the variance of the difference between their adjusted means is $(\frac{1}{3}+\frac{1}{6})\times$ error mean square, 3 and 6 being the replications of the two treatments. It would be so whatever happened to the plots assigned C, D and E, always provided that none received A or B in the confusion for that would destroy the proportionality of the two treatments.

These considerations point the considerable merits of simple designs like randomized blocks when losses or mishaps are feared. For example, if an experiment is designed like this:

Block I	A	B	C	D	E
II	A	B	C	D	E
III	A	B	C	D	E
IV	A	B	C	D	E

each treatment starts off orthogonal to blocks and proportionate to every other treatment. It would indeed require a cataclysm to destroy all those useful features.

CHAPTER 9

THE MEASUREMENT OF PERENNIAL PLANTS

". . . corroborative detail, intended to give verisimilitude to an otherwise bald and unconvincing narrative."

W. S. GILBERT, *The Mikado* Act II (Pooh Bah).

90. General Survey of the Problem

In a sense the problems of measurement underlie all others. Data that relate to the trivial cannot be worked up into information that is profound. Figures of themselves mean nothing until they are absorbed into thinking, a process that can be furthered by analysis but also hindered if statistical virtuosity is allowed to take over from insight. No system of measuring and recording, however well conceived and carried out, can be a substitute for the intimate and almost personal relationship that should subsist between an experimenter and his trees. Plants are not measured primarily to provide data for statistical analysis but to put on record and to describe to others what was apparent to the experimenter at the time. The other purposes, e.g. confirming what was suspected, possibly by use of a significance test, though important are yet secondary. One of the surest ways of judging an agronomic research team is to ask them what they are accustomed to measure and then to ask them why. Their answers will illuminate as nothing else can the level at which they have looked at the problems before them.

In the middle of the nineteenth century there came a realization of the ways in which society can be described in quantitative terms. The demonstration that cholera was transmitted by water and the identification of the sources of puerperal fever were statistical in approach and of obviously good motivation, but other applications seemed inhumane. Among others, Charles Dickens was angered and his novel, *Hard Times*, is an attack upon the sort of statistical approach that measures everything and understands nothing. His general thesis will command the respect of all good statisticians, who see their task as the promotion of understanding. However, the great novelist was mistaken in some ways and it may help in clarifying the role of measurement to consider his attack.

Like many other great works the book opens with an incident that is irrelevant apart from setting the scene. Mr. Gradgrind is inspecting a school and asking the children questions. He points to a little girl and demands to know "What is a horse?" As it happens, her father works with horses and she has been familiar with them all her young life, but she proves to be quite unable to describe one. However, a boy satisfies the critical Mr. Gradgrind with an answer that is accounted superb. A horse, he says among other things, is a graminivorous quadruped possessing forty teeth. His knowledge saves the reputation of the school; it also points Dickens's mistrust of the statistical method, because really the boy knew much less about horses than the girl did.

Whatever Dickens may have thought, there was merit in the boy's reply. The question was "What is a horse?" i.e. "What distinguishes a horse from other animals?" To anyone who had heard of horses but never seen one, the boy's reply would have been helpful. Linnaeus made quite a reputation for himself in that way. No doubt every child in the class would have recognized a horse if one had appeared, but the boy's reply constituted a record that could be comprehended and used by others. Similarly, the measurements made on experimental trees should describe them to someone who has not seen them.

However, Dickens had more to object. Mr. Gradgrind impressed upon the teacher the need for facts and maintained that children should be taught nothing else. Here some clarification is necessary. The little girl was distinctly good at facts about horses despite Mr. Gradgrind's low opinion of her. She knew very well what horses eat and what they are

like when they are ill. Such information is more difficult to impart than giving the number of teeth, but it is factual all the same. Also, it is most helpful when it can be quantified, e.g. by giving figures for the components of a good diet. Further, living things vary, so mean values are not enough but require to be supplemented by measures of dispersion. It was a weakness of the mid-nineteenth century statisticians that they believed themselves to be dealing with facts while others dealt only with fancies. Some facts are difficult to express but they need to be recorded as much as the easy ones. Further, suitable measures are devised only with insight and they require insight for their interpretation. Figures of themselves mean nothing. Prepared and considered with imagination they illuminate as well as record.

In general, records may be classified in three ways, quite apart from the characteristics recorded. First of all, they may be "complete", as when someone counts all the fruit picked off a tree or measures all the shoots, or they may be based on "samples", as when certain fruit are chemically analysed to find the sugar content of the crop as a whole. Actually, many records that appear to be complete are not really so as, for example, when an estimate is made of the colour of a crop. If that is done on the trees, the recorder notes the colour of the fruit he can see, which may be all or only some of them. If he can see only some, they will be the outside ones and therefore probably of better colour than those that are shaded. If he performs the task when the fruit are picked and in a box, he will see only those on top. Probably that layer of fruit will have been picked at one time from a single location and it may be better or worse than the rest but not necessarily representative of them. If there are more boxes the position is eased, but again there is a sampling, the method not necessarily being ideal. The point has been well made in a wider context by Cochran and Watson (1936). It may be mentioned that any experiment at the best represents a sampling procedure, for no investigator studies all existing trees (let alone all trees, past, present and future) that receive a certain treatment but only a group of them. Indeed, it would be possible to write quite a logical statistical textbook discussing experimental design only as a section of the chapter on small sample theory.

The next classification of records is into "measurement", "estimates" and "categories". By the first is meant reading the numerical scale of an instrument, whether a metre rule, a spring balance, a calorimeter or other device. By a small extension it can include weighing on a balance. By the second is meant the estimation, usually by eye, of numerical values without actually measuring them, as when recording that 20 per cent of leaves are chlorotic without actually counting. By the last is meant the placing of individuals into groups having a definite order but arbitrarily defined limits, as when the incidence of a disease is classified as slight, moderate or heavy. In making estimates it is important to define the limits of each class rather than its mid-point, i.e. 10–20 rather than 15. All are subject to hazard, even the first, as the classic text of Yule (1927) shows. A further possibility is to ask the recorder to rank what he sees, i.e. to decide which has the highest value, then the next highest and so on (Mahlstede and Lana, 1958).

The last classification is into "direct" and "indirect" records. Thus, the weight of a tree can be measured directly only at planting, transplanting or grubbing, but treatment differences can sometimes be gauged from trunk circumferences. Again, it may be too laborious to measure the length of all new shoots on a tree, but if a pruning method is intended to remove a constant proportion of new growth, the weight of prunings may well serve instead.

Plainly the best records are those that are complete and direct, but an experimenter who insists upon them may waste money or reduce recording to the point where his work suffers. The matter requires some care, both in deciding what records are really needed and then in deciding how they shall be obtained. A general warning is here required. Most experimenters record too much and do it badly. It is no good producing files of data sheets that no one will ever have time to look at properly. On the other hand, those that are to be examined must be of a standard to bear the weight of inference to be placed upon them. That may lead to a strict limitation on what can be recorded. Somewhere between the two the ideal is to be found.

91. Location and Conditions of Recording

There are two kinds of research institutes, those that record as far as possible in the field and those that bring as much as possible indoors before measuring it. Each is a little contemptuous of the other, pointing out that recording requires a lot of care and the precautions necessary to avoid error would be quite impossible if the other method were used. In fact, each approach calls for its own kind of precaution and either can be adopted providing everything, e.g. equipment, the training of staff and farm procedures, are geared to it. A decision once made is difficult to change because so much has to be changed with it. If a change is made, everything has to be thought out again from the beginning.

If recording is to be carried out in the field difficulties of identification are at a minimum. The staff will quickly learn to stack boxes of fruit under the tree from which each came and to leave prunings, etc. in tidy heaps. The trees have to be clearly labelled, which can be a disadvantage where estimates and categories are involved because the treatments cannot be kept secret. Weights have to be measured using tripods and spring balances on account of the roughness of the ground. Fruit cannot be kept temporarily in store till the recorders get round to measuring it, which can cause annoyance. Rain can increase the tare weight of boxes and bias the apparent weight of fruit. Plots must not be too small, a point readily appreciated by anyone who has tried to weigh three-days' crop from one strawberry plant in the open air. Records have to be written on tough paper in case of rain and transcribed later. The difficulties are numerous, but mistakes due to the premature removal of fruit are kept to a minimum because the farmer can usually see if the recorders have departed or not. (In any event, it is wise for them to leave a distinctively coloured label on anything they have finished with.) If mistakes of identity are made, there is often a chance of rectifying the error if the fault is discovered quickly enough. Incidentally, not even the greatest enthusiasts for indoor methods have found out how to record trunk circumferences and blossom counts in their way, so some rudimentary organization for field recording is required anyway and that is a reason for using it for everything. In general, field recording is at its best where the weather is predictable and the main records require observation of the trees themselves. Some of the early East Malling papers (Hoblyn, 1931; Irwin, 1931; Rogers, 1931) remain relevant. A later publication is that of Neumann (1958a).

Anyone who wants to record indoors, whether in a rough shed or an elaborate centre with electronic equipment, must begin by devising a good system of labels. For example, with a crop each container to be used should have its own individual number written on it. Then, when fruit is packed into boxes and before anyone removes anything from the field, someone should go round and note not only how many boxes were needed for each tree or plot but the actual numbers on them. Then, if an unidentified box is found lying around, there is some chance of repairing any harm its loss has caused. Also, the recorder can detect that he is a box short for a certain plot and can tell someone its actual number so that a search can be made. The position is eased where the pickers are paid by results and so have an incentive to keep together the fruit they have gathered until it is weighed. (The difficulty here is rather to keep each picker to his correct plot; a tendency to pick with friends or to stray where fruit is more easily come by needs to be firmly checked.) With some species, e.g. tea, that is a natural way to record. It is a help for each picker to take the label left for him on a plot and to tie it on to his basket or box for the recorder to note when weighing. If faults are noted quickly, the picker can be taken back to the field and asked where did he get the crop from, but he may not be co-operative if he has been shirking. One great advantage of centralized recording is the facility for punching data on to computer cards or tape from the start with no need for transcription later, a tedious process that can be very productive of error. In general, centralized recording is at its best for crop in conditions where the weather is unpredictable, or each plot is under the control of one picker or it is a convenience to be able to store fruit quickly after picking. It is not usually feasible with tree records as such, except for those made before planting or after grubbing.

It has already been mentioned that in a recording centre data can be made computer legible right from the start, though it is advisable to use a device like a card punch or a teletype that keeps a printed record of what is punched. In the field also various methods are available for perforating cards as measurements are made, though they can be awkward in rain and rarely provide printing (Berry, Cleaver, Nelder and Salter, 1966). From them it is a simple extension to automatic recording, in which a box of known tare weight is put on the scales, a button is pushed and computer-legible perforations are made in a card. Opinion is rather divided. Where picking goes on over a period or, as is often the case in the tropics, continuously, the mass of data to be transcribed can be enormous, whereas a computer readily performs additions and much else just by using the perforations. Also, if the apparatus breaks down and it is found at the end of the day that the perforations have not been made, the loss of data from one pick is serious but probably not disastrous. On the other hand, where all fruit is picked at one time, a failure of the automatic recorder causes the loss of a complete year's records and that cannot be tolerated in any circumstances. The situation is eased if a record is typed out whenever the button is pushed, partly because it enables a fault in the perforator to be remedied by transcription later and partly because it enables the operator to check that the data look reasonable. However, if he has to be continually thinking about the movement of fruit and answering the questions of new assistants, he will be an unusual man indeed who stops whenever something has been typed so that he can check the record against what he sees before him. Most people, after several days during which nothing excites suspicion, will just take the chatter of the typing as a sign that the record has been made and will pass on. The realization that for the latter part of the day the device has produced nothing except zeroes will come too late.

Where the recording procedure results in computer-legible cards or tape, it is most important that they are put *at once* into a program that will look for and report upon the commoner kinds of error and will work out treatment means. If haste is made, there may still be time to rectify mistakes; if not, the results may be nonsense but no one will know.

One possibility that should not be neglected is that of photographing trees from a standard distance and direction (Borgman, 1953), which is not difficult in a regularly planted orchard, and then measuring the photograph. As a means of finding the height and spread of the crown it has obvious advantages over measuring with chains and poles, while it is even better when measures like crown density are required, especially if a device can be evolved to measure the passage of light through parts of a photographic negative. It is unlikely that anyone would pay for an aerial photograph just to provide a measure of size of individual trees, but a suitable negative can nonetheless be most informative, especially in showing areas of poor growth. For most purposes black-and-white is ideal and the best time for deciduous trees is winter, but aerial colour photographs can be likewise informative about the incidence of disease. Perhaps the warning is unnecessary, but no one should be so enthusiastic about the technology as to neglect the obvious precaution of exploring the site to see what the differences look like on the ground and perhaps taking specific measurements to record what the aerial photographs have suggested.

92. Sampling Methods

The experimenter who proposes to use sampling methods need not be put off by the fear that he will necessarily introduce inaccuracies into his results. As has already been stressed, an experiment is itself a special form of sampling technique and has itself an error, which arises from the trees selected for each treatment not being entirely representative of the populations to which they belong, being in fact a sampling error.

Given a method of sampling that is random, the worst that can happen is that the error will be increased and consequently more replicates will be needed—not in itself a tragedy. Many investigators would prefer sampling from 100 trees to recording 80 in full. Dagnelie (1964) also has considered sample size.

In fact, an experiment as designed has already an error inherent in the trees (here to be written σ_i), which it would have even if recorded in full. The question is, what happens when the technical error of sampling or, as it is more usually called, the random error (here to be written σ_t) is added to it? Contrary to what might be thought, the larger σ_i is, the larger σ_t may be, i.e. if the trees are already very variable, there is no point in measuring them with extreme precision.

For the case of random sampling the two quantities can often be evaluated thus: Take an existing trial and sample each plot twice, add the two sample values together and work out an analysis of variance on their sums. The error mean square will then provide an estimate of $2\sigma_t^2+4\sigma_i^2$. Now take the differences between the duplicate pairs of values, square them, add the squares and divide by the number of plots to obtain an estimate of $2\sigma_t^2$. From these two estimates it is possible to obtain values for σ_i^2 and σ_t^2. It should be remembered that the results are themselves estimates and subject to error, from which it sometimes happens that σ_i^2 comes out negative. That can never really be so; in such instances it is best just to conclude that the technical error is large in relation to the inherent, i.e. the sampling method is a bad one. The calculation should in any event be repeated for a range of trials if reliance is to be placed on the conclusion.

It is now possible to calculate the effect of sampling on the number of replicates needed. Supposing that, with complete records, a certain trial would need R replicates to show the desired differences, then, with use of the sampling method, $R[1+(\sigma_t^2/\sigma_i^2)]$ would be needed instead. The aim should be to devise methods such that $\sigma_t < \frac{1}{3}\sigma_i$, i.e. for which the technical error of sampling is almost lost in the inherent error, thus calling for no important modification in trial design.

The above work assumes that all selections for the sample have been made at random, i.e. if the method is to take four fruits from each of three bags of similar size, first the bags must be selected at random and then the fruit from within each. In practice, that is rarely done and it is sometimes hard to decide how far a convenient method of selection will suffice One of the commonest devices is to disperse the sample, e.g. to choose three bags from eleven by taking the second, sixth and tenth in the order of picking, the aim being to take bags, or whatever it may be, at a regular spacing with, as far as possible, a half-space before the first one chosen and another after the last. Again, two branches may be selected on a tree by taking the most northerly and the most southerly.

In general, a systematic sample is rather more precise than a random one, though its technical error is not so easy of calculation. One way is to take a sample for each plot of an existing trial and then to record the plot in full. The differences between the two results, squared, added and divided by the number of plots give an estimate of the value of σ_t^2 (not $2\sigma_t^2$ as in the case given above). Now, an analysis of variance on the true values, as given by the complete recording, will obviously give an error mean square that is an estimate of σ_i^2. Another way is to carry out the method in duplicate as before using random selections at every point and then to argue that, if dispersal had been used instead, the results would have been, if anything, more favourable to the sampling method. When once σ_t and σ_i have been evaluated, the rule for increased replication is as before.

A difficulty with non-random samples, whether dispersed or not, is the risk of bias. Thus, the writer has recommended elsewhere (Pearce, 1947) that fruit size in apples should be judged from a sample divided between the tops of every third box, which leads to a bias from small apples slipping down between large ones. It is true that the extent of bias can be calculated and allowed for, but the point has to be guarded against. Hence, a better method for gauging the value of a dispersed sample is the first one given, i.e. taking a sample from each plot followed by a complete record, because that provides an easy test for bias. Unfortunately, complete records are sometimes quite impracticable even for a trial of recording techniques. It should be mentioned that a constant bias has no effect on treatment differences.

Even samples that are random at every stage can lead to bias if the first choice is between classes of unequal size. Thus, if raspberry leaves were being selected for chemical analysis by first selecting canes and then an equal number of leaves on each of those canes, the leaves on a small cane would have a better chance of being represented in the final sample than would those from a large one. If they were in some way different, the sample would be biased. Love, Cowgill and Hopp (1952) have published an interesting study on the sampling of bark. The sampling of soils and leaves in orchards has been studied by Holland, Little, Allen and Dermott (1967), leaves by Holland (1968a) and soils also (1968b). Clarke and Corke (1955, 1956) have studied disease incidence in blackcurrants.

93. Estimates and Categories

It is not easy to assess the value of records of this sort, because it depends very much upon the person making them. At the best, they provide a very easy means of showing differences; at the worst, quite large differences can be missed, either because the observer is not good enough or because he has some private anxiety which inhibits judgement. Some years ago the writer carried out some investigations into the varying abilities of different people to categorize fruit colour in apples and was surprised to observe how he himself varied from one trial to another, until he noticed that he did best when he had no qualms about the way the organization of his assistants was working out and worst when he feared a muddle. To sum up, if a person, knowing nothing of the treatments and their probable effect, detects differences, there is almost certainly something there, but a negative result may be due to personal factors.

Ideally the recorder does *not* know what the trial is expected to prove or, at all events, does not know which plots are expected to show the differences. However he reacts to such knowledge, he cannot remain unaffected by it. In practice, however, recorders cannot be changed so often and they should be taught to regard trials objectively so far as that is possible.

Supposing that the figures produced do show a significant difference, it is not always clear what has in fact been observed. Thus, estimates ostensibly of the number of blossoms may have been determined chiefly by the size of individual blossoms or the earliness of their opening. It is not unusual for observers to have their quite definite statements contradicted by subsequent measurements, but they must have detected differences of some sort.

These sardonic reflections apart, estimation and categorization are very valuable tools in the assessment of experimental results, but like everything else, they must not be used uncritically. One point that needs attention is the choice of scale. For estimation it is rarely advisable to use more than *ten* grades, and in categorization, where there is no numerical background, more than *five* is usually disastrous. They cannot be kept separate in the observer's mind and the confusion only obscures what might otherwise be plain.

The writer has a high opinion of two kinds of categories he has used in the past. In the one the recorder makes a preliminary survey of his plots to determine Grade 1, bad; Grade 3, average and Grade 5, good. With these fixed in his mind he adds Grades 2 and 4 for intermediate cases and then works over his trial, grading each plot.

In the second, based on a method originally evolved by the British Broadcasting Corporation for listener research (Silvey, 1944), he ordinarily has only three categories, 2 for "below average", 3 for "about average" and 4 for "above average". Also, for emergency use only, he has 1 for "dreadful" and 5 for "really excellent", but they are superlatives to be used carefully. The advantage of the scale lies in its assessing observers no less than experimental treatments. An excess of 1's and 5's marks down the enthusiast with little judgement, while an excess of 3's shows the timid person who will not venture an opinion. A disproportion of 2's and 4's shows the man who does not know what "average" means.

Whatever scale is chosen it should give a fairly reasonable statistical distribution of values. Thus, a mycologist categorizing lesions on fruit might be tempted to evolve a scale like: 0, no lesions; 1, one lesion; 2, two to ten lesions; 3, eleven to twenty lesions; 4, more than twenty, but when he comes to use it on a sample of 100 fruit perhaps he obtains the distribution,

26, 0, 5, 2, 69. To a statistician that is absurd, yet it often arises from scales like 0, "none"; 1, "a few"; 2, "many". In such cases it may be best just to use the "zero count", i.e. the proportion of units (i.e. fruit or leaves) without any lesions or insects or necrotic areas. It often happens that $-\log$ (*zero count*) is proportionate to the mean number of lesions, etc. per unit, which is a reason for using zero counts if the rule can be established for the particular case. The method may call for a larger sample than would otherwise be taken, but it reduces recording of each unit to the minimum.

Where several recorders are used, it is better for them to work independently and to average their results afterwards rather than to let them come to an agreed result for each plot. Averaging results in this way and using either scale avoids the danger of moderate and balanced observers, who stick to the middle grades, being overborne by dogmatists, who think everything either wonderfully good or astonishingly bad—a great risk where several observers use a scale without clear instructions.

Where a plot contains several trees, it is usually better to grade each tree separately and to average the results, though categorization of whole plots is quite valid.

Altogether, ability to categorize well is not a bad test of character because defects such as assertiveness, timidity or lack of self-confidence all militate against success. In general, experience does not count for much unless the observer has frequently checked his results against actual measurements. Indeed, a knowledge of trees can accentuate the difficulty already mentioned of having to decide what an observer actually records. Thus, an experienced horticulturist asked to estimate the crops on a series of trees might observe that one of them had about a bushel more than its neighbour, but he might go on to notice its dark leaves and extensive shoot growth and thus to exaggerate its excellence; whereas a novice, equipped only with a good idea of what a bushel of apples looks like, might estimate the difference correctly. Best of all, however, would be an experienced person who had had opportunities to check the accuracy of his past attempts, e.g. a fruit picker, whose ease of working depends in some measure on knowing exactly how many boxes to take to each tree. However, even good observers have their individual ways, as was noted by Yates and Watson (1934) with wheat.

To sum up, success in categorization and estimation depends largely upon the choice of recorder and the clearness and nature of his instructions. With respect to the first, it is best to choose someone who knows his own mind without being dogmatic, who does not know what results the experiment is expected to give, and who is either so new to the job that he will be careful or, if experienced, has been accustomed from time to time to checking his results against actual measurements. Also, he should not be colour-blind. As to the second, he should know clearly what he is supposed to be observing and should have given some thought to distinguishing it from other and similar qualities; also, he should understand clearly the scale he is using. Perhaps the worst way of selecting recorders for the job is to send them round with an expert, who discusses each plot in turn, and then to choose the candidate whose results agree most often with the expert's. As a means of selecting the timid and ingratiating it could hardly be bettered, but for the selection of recorders it has no merits at all.

94. Indirect Measurements

In efforts to reduce the recording load there is always the risk of introducing biases and distortions, especially when indirect measurements are proposed. For example, a quantity, x, can be readily found and someone proposes to use it instead of y, the quantity sought, pointing out that the two are always found to be highly correlated. There is a most insidious danger here, because treatments often break relationships that are fixed as long as trees are treated alike. Indeed, it can be said that treatments are valued precisely in so far as they enable one thing to be controlled independently of something else that is usually related to it, i.e. to obtain fruit without first growing a larger tree, or to supply nitrogen without making the soil

acid. Treatments have the same differential effect on records. For example, in general trees that grow more crop more. However, if a tree is shaded more than others it may well put out shoots to find the light and in consequence will have fewer resources with which to bear fruit. Someone may suggest in all good faith that once growth records have been obtained there is little need to measure reproductive activity as well (blossom counts, etc.) because it will necessarily depend upon size, but he could be seriously wrong if some treatments led to competition and so to shading, while others did not. Also, more severe pruning may lead to growth with a consequent loss of crop, so the suggestion could be dangerous in the context of a pruning trial also. In fact, no one should use indirect measurements without a thorough exploration of the position. Incidentally, this is another instance of the distinction between treatments that intensify differences naturally present in the environment and those that do something to the tree itself. Supplying more fertilizer leaves a tree to grow and adapt itself by its own laws and therefore is not likely to disturb relationships between the various activities of growth and reproduction, but pruning is an attempt to make the tree do something it would not otherwise have done. Obviously it must disturb at least the relationship between extension growth and trunk thickening.

Figure XXVIII illustrates this point. Crosses and circles show the values of y for a series of values of x for each of two treatments. An experimenter who tested his recording method on only one treatment might well be led to believe that x is an almost perfect measure of y, though really it might be most unreliable. In the left-hand diagram, if x is *small* and the same for both treatments, so far from the y's being equal, the treatment represented by crosses definitely gives the larger y. If, on the other hand, both treatments give the same *large* value of x, they again differ in respect of y, though in the opposite direction. Since there is no means of deciding *a priori* how the two lines in the diagram do lie in relation to one another, the position is chaotic despite the high correlation between x and y for any particular treatment.

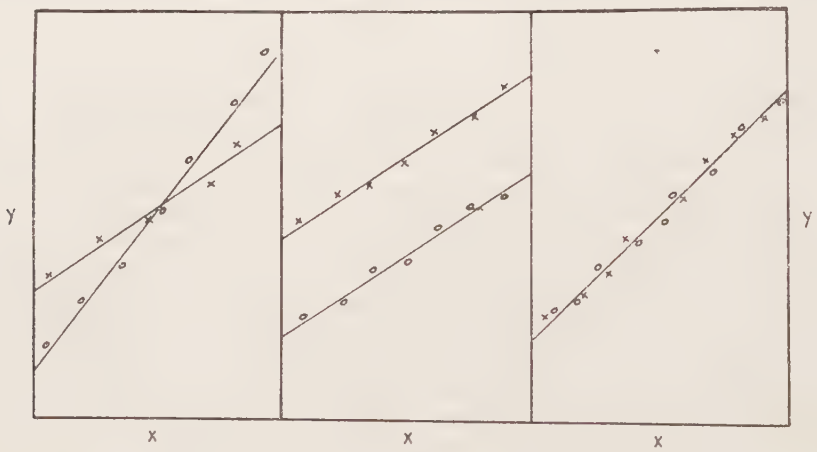

FIGURE XXVIII. Possible relationships between y, the quantity estimated, and x, its estimator.

If, as in the middle portion of Figure XXVIII, it appears that the lines are at least parallel —not an uncommon state of affairs—it may well be that use can be made of x if some information can be gained on the factors determining the distance between the lines. Ideally, however, the two lines should be identical as in the right-hand portion of the diagram, but that does not always happen.

This difficulty arises at times with almost all methods for simplifying records. To take sampling, for instance, it has been suggested that asparagus crops could be measured for only

the middle of the season, but that would unduly favour those varieties or treatments of a mid-season tendency, i.e. the relation between crop during the selected period and over the whole season may itself depend on the treatments under investigation. Again, to take estimates, though many people can accurately compare crops of trees when all are similarly treated, they easily go wrong when some of the treatments lead to the fruit being more readily seen. Even irregular picking can have that effect and Neumann (1960) has emphasized the need for care. With indirect measurements the difficulty is very common and usually requires careful investigation before a method can be recommended. The writer has published an account (Pearce, 1952a) of the position when trunk girth is used as an indirect measure of tree weight in apples, a field where the relations are especially complex. Another example occurs with coffee, where the weight of "cherry" is related to the weight of "clean coffee" in a manner that depends upon soil moisture.

It is not, of course, suggested that this objection always applies, though it should always be considered. In about 90 per cent of the published methods for the simple recording of fruit experiments it has been ignored, so some discrimination is needed before taking over a method described by another investigator.

With indirect measurements more than with any other kind of simplified record the relationship between the quantity actually written down and the true value is likely to be affected by the treatments. Thus, it is often true that the weight of leaves is closely related to the product of their length and breadth, but that would be of no use for a comparison of leaf weight between different varieties where the shape and thickness of the leaves might be expected themselves to vary.

Indirect measurements for size records have been suggested for grapes (Bessis, 1960a) and apples (Pearce, 1952a, 1956b; Holland, 1959; Moore, 1966; Holland, 1968c; Westwood and Roberts, 1970). The extent to which size records can be used instead of crop has been considered for apples by Wilcox (1937b, 1945) and Urui and Lilleland (1959). One indirect measurement of some importance is the substitution of the weight of individual fruits by their diameters or *vice versa*, because some grades use one measure and some the other. The situation for citrus has been considered by Turrell, Garber, Jones, Cooper and Young (1969).

95. Measurements of Size and Growth

Since growth may be defined as change in size these two groups of records are related. Mostly there are a number of alternatives and the experimenter must be careful not to regard them as necessarily equivalent. Thus, with most woody, dicotyledonous species the trees grow in two ways, (1) by making new shoots and extending old ones and (2) by thickening those already made. Since the application of treatments often disturbs the balance between the two, some measure is required of each. So far from the two being equivalent, any imbalance that can be detected between them could be of high importance. Again, with strawberry plants it is possible to measure both height and spread of the crown, but some are content to measure only one on the grounds that the two will necessarily tell the same story. On the contrary, the first indication of virus infection may well be a gain in spread at the expense of height as plants collapse. Similar comments apply to the height and spread of the crown of apple and pear trees, where heavy cropping will pull down the branches and alter the cropping potential for the future. It is not a question of deciding which record is "best", but of discerning the relationship between them, a task that is often multivariate in character (see Appendix II). The experimenter will usually find that there is much that various measurements have in common and an important vector may be one that combines them. Equally their imbalances and the nuances of their different accounts of what is happening may be important and other vectors will reveal the nature of these other effects. A study on these lines by Moore (1975) will repay attention and there are others (Pearce and Holland, 1960, 1961; Pearce, 1965c; Moore, 1968b).

Measuring the total length of all the shoots on a tree can be laborious but it is not impracticable. For many species it is possible to use the method of Hoblyn (1931). The recorder notes the main scaffold branches and puts them in order, the morphologically highest first. He then runs his eye up each branch, choosing the higher bud at each bifurcation until he reaches the top shoot, which he measures. He then retreats to the bifurcation below it and measures the other. He proceeds in this way right back to the trunk. Whenever he arrives back at a bifurcation from the upper shoot he takes the lower one; whenever he arrives back along the lower shoot, he retreats further to the next bifurcation, noting difficult decisions, that might have been made differently, with a chalk mark. In that way he covers the tree, main branch by main branch. If a sampling method is required, he can do one of two things. Either he can record one branch only, referring his results to the tree as a whole by use of branch girths, or he can measure only certain shoots, e.g. every tenth (Jolly and Holland, 1958). In the latter method the dispersion of the sample is much better, but if the major part of the work lies in moving ladders little may have been gained. Nevertheless, total shoot growth can be measured if it is really needed.

Measuring the total length of new growth can be exceedingly laborious on a plant of any size, though a useful measure of shoot growth is obtained by selecting certain typical shoots, e.g. those on the outside of an apple tree and at an angle to the horizontal between 15 and 75 degrees (Wilcox, 1937a) and measuring a sample of them dispersed round the tree. It does not give total shoot growth, but it provides a measure of mean growth per shoot. As has already been mentioned, some indication of total shoot growth is obtainable for spur-pruned trees from the weight of prunings, while for peaches the writer has measured every eighth shoot instead of them all. Blake and Hervey (1928) thought that some allowance should be made for there being years giving many shoots and others giving a few long ones. They evolved a "New Jersey standard" based on the number of twigs and the number of branches longer than 24 inches for use with peaches. So far as the writer is aware, no wholly satisfactory method is known for measuring shoot growth on a large plant, which is a great pity because the lack of one is serious.

In measuring shoot growth on trees that are pruned, a decision has to be taken whether to do so before or after pruning. If the aim is to measure the activity of the tree, it is better to measure before; but if an indication of future bearing capacity is being sought, it is better to do so after. Given a uniform method of pruning, it usually matters very little either way.

Measurements of thickening, unlike those of extension, are easily made. It can be done by recording the trunk girth at yearly intervals. In general, a measure of the circumference using a tape is better than one of the diameter using calipers. The place of measurement should, if possible, be at least 20 to 25 centimetres from both the graft union and the lowest branch and should be marked either with paint or nails, so that it can be identified in a year's time. The geometry of a trunk is of more interest than it appears to be (Chaturvedi, 1926; Matérn, 1956).

Because of the ease of taking trunk measurements and the difficulty of recording extension growth, there is a tendency to use only the first. As has been said, neither tells the whole story. Indeed, for trials of varieties, whether of rootstocks or scions, and of pruning methods trunk measurements alone can be disastrous. There is, however, some evidence (Pearce and Davies, 1954) that trunk shape of apples is a useful measure, because a tree that makes much growth and so forms a heavy head tends to splay out the top of the trunk. Hence, a difference between circumferences at two levels can be informative. Sometimes it is useful to measure branch girths and to combine them into a single figure by adding them and dividing the sum by the square root of their number (Pearce, 1956).

A common class of size record is illustrated by the height and spread of the crown. Height is useful in showing the difficulties of picking and spread in suggesting the planting distance required. If they must be combined to give a single measurement of size it may well be best to assume a geometric shape, e.g. Hedrick and Anthony (1919) assumed a cone. However, as

has already been mentioned, this is a case where the slightly different stories told by the two records can reveal something of importance. The measurement of height has been considered by Turrell (1946) for citrus.

Probably the best of all measures of the size of a plant is its weight, but that can only be recorded at planting, transplanting or grubbing. To consider individual species, with strawberries height and spread of the whole plant are very useful. For plants that are cut back severely annually, such as raspberries, the number and length of new shoots are almost the only records needed and, with that particular species, can sometimes be measured by counting the number of crossings of canes with supporting wires. Palms can sometimes be measured by taking the length and breadth of a sample of leaves. Rigney, Morrow and Lott (1949), working with blueberry, used the number of fruiting shoots.

Quite apart from its value as a size record, trunk circumference is worth recording every year for a special reason. If a tree dies or if it becomes scion-rooted, continuous girth records provide a means of looking back and seeing when abnormality first appeared.

One point to be noted concerns distribution. Size records represent the effect of a growth process. The application of fertilizer does not cause a sudden change in size but raises the growth rate so that after a time the tree is larger than it would otherwise have been and so it is with other treatments. Accordingly a logarithmic transformation, which changes size to growth rate, may be useful. Similarly, if growth is being studied rather than size, the difference between the logarithms of final size and initial size may be the best variate. Analyses of increments, on the other hand, are apt to be in statistical trouble from the nature and distribution of the variate itself.

96. Measurements of Fruit and Blossom

In many investigations of a practical nature measurements of crop weight are of unique importance, because the ultimate end of a fruit grower is to grow fruit. Nevertheless, other fruit records are often equally essential because some indication of quality is needed as well, though the precise nature of those records must depend upon the species and the purpose for which it is being grown. Blossom records are also of value as indicating the plant's potential crop unvitiated by factors such as poor pollination and frost, while time of blossoming is also frequently of interest for species in which interpollination of varieties is a difficulty.

In all measures of reproductive activity there is the possible complication of biennial effects. The solution usually is to insist that figures are meaningless unless based on an even number of years. That can itself raise difficulties if there are political reasons for presenting an annual report on each experiment. One possibility is to report each year total figures for the last two. At whatever cost efforts should be made to avoid reporting one year's results, because they are likely to show all differences non-significant. After repeated reports of that kind, the standing of the research institute will be undeservedly low. These remarks apply with less force to tropical crops, partly because growing seasons are usually less clear-cut and partly because there are suggestions that cyclic effects sometimes have a longer period.

In fact, the period over which crops should be accumulated raises rather complex questions. For one thing, a set of trees tends to obliterate its own cropping patterns. Those trees that cropped well over the last few years may well have grown less and therefore they may crop less well in future. Also, if they have spread their branches under the load of crop, they may develop quite differently from those that earlier cropped less well and instead built up a good frame of branches. These remarks have some relevance to calibration and covariance (see Chapter 4). A related consideration is that of dividing the life of a long-term trial into periods for study. The report of a fertilizer trial (Greenham, 1965) illustrates the point well. Treatments that are effective in establishment do not necessarily have much effect later, while a treatment that aids fruit setting may do little in early years. If accumulated crops are presented, the results can look very complicated indeed, while useful information may be lost if it is not made apparent at what stage of development a beneficial treatment had its effect.

When it comes to the actual mechanics of recording a crop, whether in the field or in a recording centre, it is often sensible to adopt a sampling method based upon containers. The number of full containers, whether per crop or per tree, is counted, perhaps to the nearest quarter container, and a sampling method is used to obtain a mean weight by which to multiply the count and make it into a weight. At this point a decision has to be made. If the mean weight per container is to be obtained from a sample spread over the whole experiment, the investigator should be clear that what he is really presenting is a volume record expressed in terms of weight. If a certain treatment gives fruit of a different specific gravity or of different size or shape so that it packs differently into the containers, results will be biased and it would have been better to have left the volume record in its original units. An alternative is to take a sample weight per container for each treatment separately, but to do so is to invite bias of a different sort. If for a particular treatment a high value is obtained, it will be applied to each plot of that treatment and will appear in the treatment mean without having been allowed for in the estimation of the error variance, so standard errors and significance tests will be wrong. The only safe course is to weigh one or more containers from each tree or plot and to make a separate estimate for each such unit. At this point the recorders will probably complain that a large part of their work lies in either setting up their weighing apparatus at each plot (if they are weighing in the field) or in labelling (if they are based in a recording centre) and that they are not being helped. They are, however, being accorded a little relief and they should accept that crop records do need a lot of work.

The accuracy of this method depends upon the constancy of weight of fruit in a container, for obviously, if all containers were identical in what they held, the method would be perfect. Actually it is necessary to estimate two variabilities: σ_p, which represents the variation *from one plot to another* of the weight of fruit in a container, as measured by its standard error, and σ_c, which is the standard error of the weight of contents between containers *from the same plot*. To evaluate these quantities, it is best to take an existing experiment and to select two containers per plot at random and to weigh the contents of each. Then σ_p and σ_c can be found using the same method as for σ_i and σ_t in Section 92. However, Szczepański and Kłossowski (1963) found sampling unsatisfactory for apple crops and preferred full recording.

With crops that are picked frequently an obvious possibility is to record only certain harvests. One advantage is the chance of relaxing standards on picks that neither precede nor follow a harvest selected for recording, which does not apply if alternate harvests are studied. The problems have been well studied for tea (Eden, 1931; Wellensiek, 1938; Eden, 1941; Visser, 1958). Valuable studies on the incidence of crop have been made by Cooke (1962) and Vernon (1969).

When working out a technique for future use, containers should be chosen truly at random; but when the method is actually being applied there is no objection to dispersing them systematically over the plot, either according to the order of filling or so as to include fruit from as many trees as possible. Such dispersal should prove more convenient and might improve precision.

However, fertility expresses itself in various ways, as Bessis (1960b) pointed out for grapes.

Another crop record that is often of interest concerns mean fruit size. It may be made by taking a sample and then either grading each individual fruit or weighing the sample and thus evaluating the mean weight of 100 fruits. The individual records are of greater value because they show the proportion of crop unsaleable on account of size, but the greater labour of taking such detailed records makes the "weight of 100 fruits" of greater general utility. At East Malling sampling methods have been described for apple, pear and plum (Pearce, 1947) and cherries (Taylor, 1951a) depending on the dispersal of the sample between the tops of containers. For apples there is a tendency for the smaller fruits to slip down and miss the sample, so correction by regression is needed to allow for this. On the tree larger apple fruits are found on the outside than on the inside.

Fruit colour is a quantity that needs careful definition before it can be measured. It may mean proportion of surface flushed, the hue of the flush or the hue of the ground-colour. For the proportion of apple surface flushed, it has been found (Pearce, 1947) that there is a strong tendency for the fruit to lie in the boxes in layers of similar coloration, so an effective sampling method required the taking of fruit from a vertical column in each box to disperse it between the layers. The method is to form a hole in one corner of the box, one hand removing fruit while the other holds back the "wall" thus formed. Sometimes it is necessary to take a firm grip on the last fruits to be removed and to pull, thus collapsing the wall, but usually both hands can be withdrawn at the end leaving a well-defined even hole, a quadrant of a circle in cross-section, like a well. The chief skills required are a certain delicacy of touch in the finger-tips and an ability to judge how wide a hole is needed to yield a specific number of fruit, both of which come with practice. The method works well for apples, less well for pears because of their shape and not at all for cherries because of their stalks. Each fruit of the sample is then judged to fall into one of four grades (1, up to a quarter flushed; 2, a quarter to a half; and so on) and a mean grade given. At least 60 apples are needed for each plot, unless the variety is one with a clearly defined flush. Usually, however, the amount of flush can be estimated by eye, as can the two hues; though it is far from easy to judge one without interference from the other two. If East Malling experience is anything to go by, such estimation should always be done on the tree, because the tendency of fruit to form layers of similar colour in the boxes makes unreliable an estimate based on the top layers only, i.e. the ones that can be seen. Of course, in a recording centre such records can be made well and easily.

Many other records on fruits are sometimes needed, mostly of ripeness and chemical content. The first resembles colour in requiring definition. Does it mean softness of tissue, sugar content, or is it just another name for ground-colour? The point is important, for the three are only roughly related and a difference in one does not demonstrate a difference in either of the others. Records of chemical content plainly must be based on samples and, in the absence of precise knowledge of variations in chemical elements throughout a tree, the samples should be dispersed as much as possible. Perhaps the safest way, when once the fruit is picked, is to take a vertical section as described for fruit colour and then, if that is too large a sample, to select every fifth (or every tenth or every twentieth) fruit in order of withdrawal from the boxes to form a sub-sample for analysis.

The incidence of fruit drop has been studied statistically by Llewelyn (1968).

Records of blossoming are mostly of two types: (1) amount of blossom and (2) season of blossoming. The first, amount of blossom, is usefully made by estimation, 0 representing no blossom, 5 the maximum for a tree of that size, and 1–4 proportionate to 5. For some species, however, it is possible to work along each branch in succession counting inflorescences—a task that is facilitated by wearing spectacles of a colour that will pass light from the blossoms while cutting out that from the leaves.

With some species, e.g. coconuts, the number of female flowers is of importance. With almost any species there are investigations that require records of blossoms set and fruit fall.

Season of blossoming may be recorded by visiting the trees every day and noting (a) date of first blossom open, (b) date of full blossom, i.e. 50 per cent open, and (c) date of last blossom. For pollination studies some such record is essential. For other purposes, e.g. the comparison of treatment effects in one year, it will usually suffice to visit the trees once and to estimate the proportion of blossoms open on that occasion.

97. Records of Flavour and Consumer Acceptability

It is of course important that fruit intended for sale to the public shall look well and be approved for flavour, but the required assessments are hard to make and usually expensive. Above all, however, it is necessary to be quite clear what information is being sought.

If the object is indeed to find out what the public like, the only method is to ask a representative sample. The most foolish method is to ask a panel of experts, who from the very fact of being expert are most unrepresentative. Such panels usually adopt unrealistic standards and an esoteric vocabulary. Categories such as "old boots" and "horsy" will serve as examples of their more pejorative epithets. For all anyone knows, the public greatly enjoy a flavour of old boots, their pleasure therein being exceeded only by their ecstacy at the flavour of horses. At least it would be as well to find out before condemning a variety or a cultural practice because it had led to such a description being applied. A representative panel would be afforded by a group of housewives. Some research institutes are quite used to entertaining parties of women's organizations and sometimes find it difficult to think of things that would interest them. An impromptu tasting trial followed by a talk on fruit quality might well be tried. Also, many research organizations exhibit at agricultural shows and fairs and it seems a pity that so few opportunities are seized of consulting the consumer about what he likes. Incidentally, it is possible to attach too much importance to flavour. The housewife buys at market or in a shop what looks good, but until she gets it home she will not know if it also tastes good.

To argue thus is not to minimize the value of specialist tasting panels, but they should be used to detect defined differences, e.g. spray taint, or to relate variation in flavour to variation in field or storage practice. Thus, there might be questions as to what effect prolonged storage or large nitrogenous applications had on flavour, not in the sense of asking whether the taste remained acceptable but of enquiring precisely in what respect it was altered. However, such a panel does not need an expert knowledge of fruit growing. Indeed, it may have to be protected against its own expertise by giving code letters to varieties or treatments to suppress their real names. Fundamentally there are four flavours, corresponding to four different kinds of taste buds, namely, sweetness, acidity, bitterness and saltiness, so a panel needs to be selected on the basis of its members' ability to detect the level of these basic flavours, both alone and in combination, a task that can be accomplished only by the use of pure chemicals (Freeman, 1956). However, although those four represent flavour in the strict sense, there are other characteristics that can be similarly identified and contribute to "taste", using the work in its ordinary sense, e.g. texture, juiciness and skin thickness. Also, odour and flavour are not always easily distinguished. Thus, old boots and horses must surely be odours, because few people can ever have tasted either.

Once the investigator has identified the characteristic that his treatments may be affecting, he can select a panel to detect it. Questions of blind-folding the assessors or of letting them see or smell the fruit are best settled on common-sense grounds once their task is defined, as is the sort of light in which they are to work.

A more difficult problem relates to sampling. It is not good enough to take a fruit, cut it into four parts and let each of four panellists have a piece; each must have part of a separate fruit, but how is it to be selected? Ideally it should be taken at random from all suitable fruit, but that calls for a large panel since one member will mark down a treatment because he happens to get an unripe specimen while another condemns a different treatment because he happens to get one that was overripe. It is true that fruit can be graded within fine limits to minimize such discrepancies, but it is then necessary to ensure that the selected grade is typical. Thus, if a storage treatment makes some fruit ripen quickly with an unpleasant odour a selection of "typical" fruit in which the defect did not occur would obviously be misleading. Also, panellists tire if asked to do too much at a time, a situation often dealt with by the use of non-orthogonal designs (see Sections 30 and 31), each panellist at one session forming a block.

However things are organized, assessments of this sort are costly. It is true that any one member of a tasting panel is engaged on the work for only a short period, but by the time the organizer has rung up to request attendance, the panellist has interrupted his usual work, has walked over to the testing laboratory, has waited a few minutes for attention, has undertaken his assessments and made his report and has walked back, a lot of time has been spent.

In aggregate, over the whole panel it has cost a great deal, quite apart from the irritation of more senior staff at the continual disappearance of juniors credited with sensitive taste buds. The cost is not necessarily prohibitive but it is necessarily high.

98. The Importance of Good Recording

In a series of experiments on a range of subjects many other kinds of record may be needed, while every species presents its own problems. In particular, records of diseases and pests are often essential and are sometimes difficult to obtain. In this chapter an attempt has been made to give examples and indicate principles but no more. Every experimenter must evolve for himself the catalogue of measurements required, a task that can be accomplished properly only if he sees clearly what he needs to know. Any research institute tends to have a list of records that it takes as a matter of course, as if a variety trial required the same procedures as one on the control of pests. Sooner or later will come the day of reckoning. It may even come because the records are no longer good enough for the purpose they were originally intended to serve. Thus, records of total weight of crop can become meaningless if the market develops a strong preference for fruit of a certain size or colour or skin finish. Records of mite infestation become irrelevant when it can be taken for granted that mites can be eradicated, if need arises and the questions come to concern spray damage. One of the chief needs in making records is to ensure that all are apposite to the problem.

The other is that they make good sense. One kind of nonsense arises when complicated measures are derived though it would have been better to have kept to what is actually observable. Coefficients of this and indices of that are necessarily functions of what was actually measured. In almost every case it would have been better to have made the measurement and regarded it for what it was, frankly recognizing any imperfections, instead of disguising it as some function of itself with a high-sounding alias that almost removes it from criticism. Similarly, when several related measurements are made, they are best studied individually and combined only in ways that reveal imbalances or estimate something intelligible, like proportion of blossoms damaged by frost. Complicated weighted means in which they are combined in some arbitrary way depending on one individual's idea of relative importance have little value. If some spray substances reduce the number of lesions by killing spores before they germinate, while others keep down the size of lesions without reducing their number, separate study of number and mean size will show the position and perhaps lead to a useful train of thought, whereas an "index of fungal activity" that combines the two measures and compounds two distinct phenomena into a single muddled idea can lead only to a confusion that will not be resolved until the original measurements are seen each in its own right. Recording requires much thought. First, the experimenter must think what he wants and then think how to achieve it. Much the same comment applies to any stage of an experiment, whether design, data analysis, the choice of treatments or the manner in which they are to be applied. Given a clear appreciation of objectives and a readiness to subordinate everything to them, anyone can carry out a good experiment. Without those requirements, no one can.

Appendix I

THE USE OF COMPUTERS IN AGRONOMIC RESEARCH

These days computers are very generally available and any agronomic research worker should know enough about them to make effective use of the facilities open to him. It should first be said that the uncritical use of computers does immense harm. When the calculations that arose from a body of data took a long time to perform, the experimenter approached his task circumspectly, recognizing it as the culmination of a long process in which the investigation had been conceived, a trial had been planned and initiated, much detailed data had been recorded and later summarized, until at last he was standing before the climax in which answers, perhaps important ones, might be found to the questions posed. Consequently, he considered carefully what should be analysed, how the data should be transformed, if at all, what partitions should be made in the treatment sum of squares and what could legitimately be included in error. Nowadays data can be thrown at the computer in the hope that something will thereby be revealed. Some believe, the present writer among them, that the development of computers has had a seriously deleterious effect upon statistical science by providing means for producing the outward shell of an analysis without the inward spirit that gives it life. However that may be, the computer is here to stay and everyone should study how to make the best use of it.

There are indeed many things that it can most usefully do. In the sphere of information retrieval, for example, it is a tool of unparalleled value. Thus, the Record Office of the Statistics Section at East Malling has evolved two retrieval systems to perform tasks that would previously have not been feasible. With one it is possible to enquire what experiments have been performed at the Station in the past on a given species or variety of a species, studying given sorts of treatments, the person responsible being a given research worker, at a given period on a given piece of land. All or any of these features can be specified in the enquiry. With the other, it is possible to give the corners of a proposed experimental area in National Grid references and to retrieve what is known about the land concerned with respect to former experiments and field boundaries, etc., including the evidence of old maps, some of which go back to the seventeenth century. The topic here, however, is rather the use of the computer in data analysis.

To start with the elements, a computer installation has two parts, the hardware and the software. The hardware is what arrives in cases, is carried indoors and put together. Of itself it does hardly anything; the software, which comprises the instructions given to the hardware, ensures its proper functioning. Neither should be disregarded. No one should say: "Now we have a computer, we shall be able to perform all our calculations" until he has assured himself that all necessary software is going to be available, but neither should anyone say: "Someone has let us have an excellent program, so we shall be all right" until he has assured himself that the software will fit that particular installation. Different configurations of hardware can affect very much the uses to which the computer can be put. Just as some people have good memories so some computers have extensive stores for information, and just as some people can think clearly so some computers have ample core space in which to perform calculations. In general, statistical analyses involve quite small amounts of data to which a great deal is done—at least, relative to many other calculations—and it is not reasonable to expect a computer installation designed to work out fertilizer subsidies to perform elaborate analyses. From the nature of its primary task the hardware will comprise extensive stores, probably on magnetic tape, for lists of farmers and their holdings, but there will be little core, since the expected calculations are quite simple.

The important thing then is to work in close collaboration with the manager of the computer installation, listening to what he has to say and heeding his advice. It is not just a

question of what hardware exists but the load on each item. For example, supposing the alternative to be available, it may be more considerate to store results and to read them off at night than to require a line printer at peak hours, or the reverse may hold. Again, some programs are written assuming a large amount of core, using it even for data, with a view to saving time by avoiding the slow business of continually dipping into stores; others use the minimum of core at the expense of time. Either extreme could cause problems on a particular installation and the manager might even offer programming help to secure ease of running. The danger is that the novice, whose program has run well at one place but is encountering resistance at another, may suspect obstructiveness when in reality everyone is trying to be helpful. Incidentally, most computer installations are run with the idea that speedy output is essential. For many purposes it is, but hardly for analyses of experiments that have taken several years to conduct. Usually it is best to accept the prevailing speed of operation, but sometimes a more leisurely schedule will secure co-operation that will not be forthcoming during a rush.

Software exists at several levels, some of it so deep in the system that the ordinary user is unaware of its existence. For example, the hardware of itself cannot perform arithmetic operations and special software is needed to enable it to do so. Indeed, it is possible to envisage a computer intended for a task like literary analysis that had no arithmetic skills at all, fundamental though they may seem to many users. Mostly, however, the skills exist and are manifested by an esoteric set of orders called "machine code", which will be peculiar to that particular sort of computer. At this point the would-be user may find himself in conflict with the programmers, who sometimes insist that no program is ever really efficient unless written in the machine code of the computer it is proposed to use. No doubt they are right according to their own highest standards, but before the novice attempts to comprehend the incomprehensible he should be told firmly to do what other ordinary mortals do and learn a computer language.

For statistical purposes the most widely used is FORTRAN (FORmula TRANslation), which is a good general-purpose language that unfortunately exists in many dialects. Its input and output procedures are extremely flexible and therefore complex. Their existence, however, is a reminder that the carrying out of calculations is not the whole, but data have to be read in and results extracted, so the range of FORTRAN is most valuable. If it seems forbidding, the beginner does not have to learn all the options at once. Some advocate the preliminary learning of BASIC, which is easier and essentially similar. Another common language is ALGOL (ALGOrithmic Language), much beloved by universities on account of the excellence with which it shows the precise form of an algorithm but less liked by practical men because of the weakness of its input and output procedures. Often installations are forced to devise their own supplementations with the result that ALGOL programs can be "installation-dependent" and will not transfer without substantial rewriting. Another common language is COBOL, which has many commercial uses but is less suited to scientific work, while PL1, which combines many features of FORTRAN and ALGOL, may replace them in the future.

The choice of language must depend upon the installation because at this point a further piece of software is needed, namely, the compiler, which translates what has been written in the computer language into the machine code of the computer. Hence, it is peculiar to the combination of one language and one sort of computer. Sometimes no suitable compiler will be available; sometimes there will be several. For example, there could be one for trying out new pieces of program. It will translate quickly, though with little regard to the efficiency of the final result, and it will have ample diagnostics, i.e. messages sent to output to specify any faults that have prevented compilation. A different one may be used when faults have been rectified. It should produce, possibly slowly and painstakingly, an efficient translation but it may well have few diagnostics. (Incidentally, compilers can take up a lot of space in the computer, which is one reason for the existence of dialects. A very small computer may be

able to carry a compiler for BASIC but not for FORTRAN II, while a larger one may be able to compile FORTRAN II but not FORTRAN IV.) For various reasons it is best to use a compiler with which the local staff are familiar.

With a large computer there may well be a higher-order language that translates very general instructions into a base language, such as FORTRAN or ALGOL, for which a compiler exists. Thus, there could be an instruction CORRELATE with two location references, which would cause the computer to correlate the two sets of data indicated; much more elaborate procedures are feasible.

All this assumes that the investigator is willing to write his own programs and there are good reasons why he should, chief among them being the reason that only he knows what he wants. If the calculation is simple arithmetically but includes some awkward special cases, he might very well try. On the other hand, it may be argued that a well-established team of statisticians engaged on similar work must have written programs for themselves that could be of use to others less experienced. There is some force in the argument, always provided that those who take over the programs understand the thinking of those who have written them, but often the user is disconcerted. For example, an agronomist in a country that lacked computers sent his data away for analysis and complained that the output he received was obviously wrong because the treatment means were not those he had worked out in a preliminary study. In fact, one of the interactions was confounded and the sophisticated author of the program had argued that in such a case, where an interaction is assumed not to exist, it is only reasonable to adjust the means to make it equal zero, an explanation that the less sophisticated user found unsatisfactory even when it was explained to him. Again, a very general algorithm, like the one in Appendix III, may deal with missing plots by just ignoring them and working out an analysis for the design that is left, thus producing a result rather different from that given by the more usual calculation of missing plot values. All in all, taking over someone else's program can be an unnerving business.

An alternative is to adopt the statistical package offered with the computer. The trouble here is that such programs are often written by people who know a lot about software and who perform prodigies of compression but appear to know little about the calculation in practice. For example, field data are rarely analysed as they stand, e.g. it may be necessary to add all harvests to express the marketable crop as a percentage of the total and possibly to transform to angles. Has the programmer thought of that? (Actually the variate control part of a good analysis program may be longer than the algorithm, because there are so many possibilities.) Again, has any provision been made for missing values? —or for covariance adjustments? —or for confounding? —or for the merging of certain higher order interactions with error? It is surprising how many packages are deficient in procedures that seem quite essential to the practical man, so it is as well to ask some pertinent questions before relying on one of them. All things considered, there is much to be said for each man writing the programs he needs. They do not have to be of complete generality, only wide enough for what he has in mind, and he will at least understand the output that they produce.

Finally, whether the experimenter is writing his own program or using someone else's, there will come a time when he needs a test, either to confirm that he has got things right himself or to discover what course an unknown programmer has taken in a difficult situation. One possibility, of course, is to use a body of data that has already been analysed in other ways, a procedure that can be ideal so long as it is remembered that computers have different ideas from people about rounding off, e.g. they do not have to round off so often because they do not have to write down intermediate steps, but, on the other hand, they do not automatically extend themselves when confronted with a long number. Consequently, queer discrepancies can arise in the last decimal place or two and it is sometimes hard to judge whether a program has a small fault or not. Two devices can be tried. (i) One is to write down a body of data that has a value equal to "one" and the rest to "zero" and to work out the analysis on paper, which should be very easy, and then on the computer. Since most

output will consist of zeros or simple fractions and since the whole exercise can be performed several times with the "one" in different places, checking is easy. Sometimes it is better to set a value equal to "one" and another, either in the same block or the same treatment or in neither, to "minus one", the rest being set equal to "zero". (ii) The other device is to construct a variate without error. For example, in a block design it is sufficient to assign to each plot the sum of the block number and the treatment number or, in a row-and-column design, the sum of the numbers for the row, the column and the treatment, and similarly in more complicated cases. If the program gives a non-zero error sum of squares, either the algorithm is wrong or the user has misunderstood the instructions.

The above may seem to imply that the only analyses of data are analyses of variance, but that is not the intention. Similar remarks apply to correlation and regression analysis and to the use of multivariate methods and much else.

Knowledge of computers is growing fast and soon this Appendix may be quite unnecessary. At least, there is no need to be afraid of the brutes, which are really very docile. When investigated, computer difficulties are almost always of one of two kinds. Either the preparation of the input has been careless or else the program is deficient or open to misunderstanding. Those who punch data have to be extremely accurate and careful. Given a figure in the wrong column or a full stop instead of a comma, the computer, which only does what it is told and has no intelligence and no initiative, will produce nonsense. Thus, if the program requires that the number of blocks and the number of treatments shall follow one another on the same line, it is no good trying a little individual variation and putting them on successive lines. Further, if such a mistake has been made, it is no good saying "Just tell the computer that . . .", because telling a computer anything involves editing the program and compiling it again, a task out of all proportion to getting the input right in the first place. Even with the greatest care, errors do occur in data preparation, so a good program will pick them out and report them instead of just stopping or going on to produce results that are clearly wrong.

Indeed, success with computers depends primarily upon the organization of a good data-preparation service and a thorough understanding of the programs and other software available. If attention can be paid to these matters and if good relations can be established with the staff of the installation, most people who try to use computers are gratified by their success.

(Despite the warnings about using other people's software, the Statistics Section of the East Malling Research Station have begun to write programs for the use of agronomic research workers in developing countries who have access only to small computers and limited ancillary services. The first is NOBLE (Non-Orthogonal BLock Experiments), written in simple FORTRAN by Mr. K. J. Martin. It is for carrying out the analysis of variance for any block design using the algorithm presented in Appendix III. If unable to perform the calculation, it can in many instances explain what has gone wrong. The variate control facilities for transformations, etc. are fairly complete. Workers in developing countries are welcome to copies.)

Appendix II

MULTIVARIATE METHODS IN AGRONOMIC RESEARCH

These days it is fashionable as well as reasonable to consider methods that use several variates simultaneously, assigning equal status to each. After twenty years' experience of the approach the writer believes that they have a useful place, though one that should not be exaggerated in importance. The trouble is that they remain inflexible and there is, therefore, a tendency to force the problem into the mould of the technique rather than to adapt the technique to deal with the problem. However, when there is no conflict, the approach can be rewarding. Some years ago the author wrote an expository article for *Experimental Agriculture* (Pearce, 1969) and he is most grateful to the Editors of that journal for permission to reproduce it here because his general attitude at that time has been confirmed rather than modified by subsequent events. He would also take the opportunity of belatedly acknowledging the help of his son, who criticized the text on grounds of intelligibility and helped to make it more clear.

MULTIVARIATE TECHNIQUES OF USE IN BIOLOGICAL RESEARCH

By S. C. Pearce

SUMMARY

Multivariate statistical methods are used increasingly in biological research to investigate the responses of organisms considered as a whole, whereas established statistical methods are usually concerned with measured characteristics considered one at a time. Multivariate techniques are mostly explained in terms of matrix algebra, which is a way of dealing with groups of numbers rather than individual ones. A brief description is given of some elementary results of matrix algebra and a method is presented whereby hypotheses can be generated about interrelations within an organism. Two techniques, principal component analysis and canonical analysis, are described in greater detail. It is emphasized that hypotheses need to be tested even though they have been generated by objective statistical means.

The rise of multivariate statistical techniques in subjects such as psychology and sociology has led to their being tried in biology also (e.g. Goodall, 1954; Blackith, 1957; Blackith *et al.*, 1963; Jeffers and Black, 1963; Hattemer, 1966; Wâlloff, 1966; Yarranton, 1967; Holland, 1968d) thus opening new vistas of statistical opportunity. It is true that their advent has not been universally welcomed, partly because they are very unlike recognized statistical methods in what they attempt to do and partly because they are usually explained in terms of matrix algebra, which to many biologists is a foreign language. In this paper an attempt is made to remove both difficulties, first by urging the relevance of these techniques to much biological research, then by explaining the concepts of matrix algebra in simple terms, and finally by considering some practical examples.

The relevance of multivariate methods

Most statistical techniques concentrate on one variate at a time. As a result, when reporting an experiment, germination counts might be presented in one table, incidence of disease in another and yields in yet a third, but when agriculturists speak among themselves they do not fragment problems in this way but try to see their plants and animals as a whole. Thus, the fact that one treatment leads to early germination may itself be the cause of increased exposure to infection and also of higher yield, even though the disease would ordinarily lead to a depression of crop. In fact, many papers contain a section entitled "Discussion", which is in essence the integration of phenomena that ideally should never have been separated,

and which perhaps would not have been if statisticians were able to consider several variates together.

It may be objected that some established statistical techniques, e.g. multiple regression, do in fact employ several variates in one analysis, but only to cast light upon one of them, namely, the so-called dependent variate. Further, this emphasis on one variate among several can lead to spuriously simple results. For example, a plant that grows more on account of being etiolated will probably have fewer blossoms, whereas one that grows more because it is well fertilized will probably have more. A regression analysis gives a relationship for a certain set of circumstances, but, whatever the result, it may be an oversimplification and can well be contradicted by the next body of data.

The development of multivariate methods is therefore to be welcomed, indicating as it does that the statisticians are at last trying to quantify the way biologists really think. It is not surprising that these new methods have had their chief success in branches of biology where the problems clearly lie in the integration of numerous related observations, e.g. taxonomy and ecology. However, in so far as all biological investigations contain such an element, it is to be expected that multivariate methods will progress further. Statisticians are still not very good at seeing organisms as a whole, but they have now made a start and their efforts call for encouragement and understanding. As it is, some biologists feel that statistics are mathematical and therefore remote from their real thoughts; the danger is that techniques using even more fearsome mathematics will be supposed to be even more remote. In intention at least, these new methods are closer to the heart of the biological thinking than the old ones.

Matrices and some properties of matrices

Since the object of multivariate analysis is to consider quantities together, it is not surprising that the methods are generally explained using matrix algebra, which is a way of thinking about groups of numbers. Scientists are familiar with algebraic quantities, each traditionally represented by an italic letter, P, x, β, etc.; in matrix algebra the subject is a collection of quantities represented by a bold-face letter, $\mathbf{P}$, $\mathbf{x}$, $\boldsymbol{\beta}$, etc. Actually, a matrix is nothing other than a two-way table of quantities. Where a scientific writer might say "Consider the values of R set out in the table", a mathematician would say "Consider the elements of matrix $\mathbf{R}$", but the principle is the same.

To attempt a formal definition, *a matrix is a set of quantities written in rows and columns, the rows and columns being meaningful, with no voids.* There are two things to be noted here. The rows and columns must not be arbitrary; the newspaper may present the numbers of the fifteen winning lottery tickets in three rows and five columns but this does not constitute a matrix because the rows and columns have no meaning. Also, all elements must be present. A table may have blanks, but a matrix by definition must be complete. (Incidentally, a matrix with only one row or only one column is more usually called a "vector".)

Most matrices used in multivariate analysis are squares, in which rows and columns have the same meanings, and most are "symmetrical". Such matrices are familiar to users of buses, which commonly display a fare table in triangular format, like this:

Town Hall			
5	Railway Station		
7	4	Red Lion	
10	6	4	St. Luke's Church

In full, this would be a 4×4 matrix with a row and a column for each stopping place. However, the "leading diagonal" from top left to bottom right has been suppressed "for triviality"; anyone can see that it costs nothing to go from A to A. Also the elements above the leading diagonal have been suppressed "for symmetry", because everyone will assume correctly that it costs as much to go from A to B as from B to A. In a mathematical context also the upper

triangle of a symmetrical matrix is commonly suppressed, but it is usual to retain the leading diagonal, even when it is trivial, in order to make clear the full size of the matrix. (Thus, in the example given, it is not immediately apparent whether the full matrix is of size 3×3 with the leading diagonal retained or 4×4 with it suppressed.) The sum of elements on the leading diagonal is called the "trace" of the matrix.

There is also the possibility of a symmetrical matrix being "singular", like this one:

$$\begin{pmatrix} 4 & 7 & -1 \\ 7 & 8 & 11 \\ -1 & 11 & -38 \end{pmatrix}$$

The word (singular) does not imply that the matrix exists alone but that it has some odd properties. To be specific about the one of importance in multivariate analysis, its elements exhibit "dependence", which here rests upon the multipliers 5, -3 and -1; if in any row or column the elements are a, b and c, then $5a-3b-c=0$.

To a mathematician singularity is an important property. To him any square matrix can, for many purposes, be summed up in a single "determinental value", derived in a way that need not be described here. If a matrix is singular, i.e. its determinental value is zero, the matrix occupies somewhat the same position relative to other matrices that zero occupies relative to other numbers.

It might be conjectured that singularity was an unusual property, but matrices exhibiting it are easily generated. In fact, if **A** and **B** are both symmetrical matrices of size $p \times p$, there will be p values of λ that make $(\mathbf{A}-\lambda\mathbf{B})$ singular. They are known as "latent roots" or "eigen-values" and their representation by the Greek letter lambda is a usual convention. The corresponding multipliers, which show the dependence of the elements $(\mathbf{A}-\lambda\mathbf{B})$, are known as the "characteristic vector" or "eigen-vector". Thus, if

$$\mathbf{A} = \begin{pmatrix} 12 & 5 & 13 \\ 5 & 13 & 4 \\ 13 & 4 & 21 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

there should be three values of λ that make $(\mathbf{A}-\lambda\mathbf{B})$ singular and this is in fact the case. One of them is 21, because

$$21\mathbf{B} = \begin{pmatrix} 21 & 21 & 21 \\ 21 & 42 & 21 \\ 21 & 21 & 42 \end{pmatrix}$$

and so

$$\mathbf{A}-21\mathbf{B} = \begin{pmatrix} -9 & -16 & -8 \\ -16 & -29 & -17 \\ -8 & -17 & -21 \end{pmatrix}$$

which has the characteristic vector

$$(8 \quad -5 \quad 1)$$

i.e. 8 (first element) -5 (second element) $+$ (third element) $= 0$. The other two latent roots are 6 and 7, with characteristic vectors

$$(1 \quad -1 \quad -1) \quad \text{and} \quad (4 \quad 1 \quad -3)$$

respectively.

The use of singular matrices in multivariate analysis

The above theorem is the basis of several multivariate techniques, in which the matrices **A** and **B** represent the variability of a set of variates. (In fact, the elements on the leading

diagonal would represent the variances of the respective variates, while the others give the covariances of the variates taken in pairs.) If **B** represents the variability that may reasonably be attributed to identified sources and **A** the variability actually observed, making $(\mathbf{A}-\lambda\mathbf{B})$ singular may be regarded as subtracting **B** from **A** as often as may be, i.e. until the determinental value is zero. Any dependence between the elements of what is left, i.e. $\mathbf{A}-\lambda\mathbf{B}$, is then studied in the hope that it will reveal the mode of action of sources of variation as yet unidentified.

To do this may well provoke protest from the mathematicians, who also have their reservations about multivariate methods. Statisticians in training are always told that their techniques should serve one of two purposes; either they should *test* a hypothesis (as when a significance level is used to answer a question like "Does this treatment affect the crop?") or they should *estimate* a quantity in a hypothesis (as when answering a question like "What increase in crop may be expected from the application of a further hundredweight per acre of sulphate of ammonia?"). Multivariate methods mostly serve neither purpose, but are used to *generate* hypotheses. Some mathematical statisticians have been quick to point out that this is a speculative and uncertain process, a fact that their colleagues in the experimental sciences have known for a very long time. It is nevertheless an essential step in scientific progress.

One important consideration concerns the actual computations, which are indeed laborious. In the old days that factor impeded the use of multivariate methods, but nowadays electronic computers are so readily available that the difficulty is reversed. There is, instead, a real danger of offering large bodies of data for processing by elaborate programs, which produce impressive but possibly meaningless results. A biological research worker needs to be on his guard on two counts. First, he must insist on knowing what the computer is doing. Matters of matrix algebra may be left to the experts, but the investigator must know what **A** and **B** represent and should be free to object if they are not what he wants. (Thus, **A** may be the variation including or excluding the effects of certain treatments. Variates may have been presented as they stand or after transformation to logarithms or percentages. The derivation of **B** can raise even more points of principle.) Secondly, he should not have an exaggerated respect for a hypothesis because it has been generated by a process that he only partly understands. Hypotheses call for consideration and test; experience shows that those generated by statistical methods are often interesting and provocative, but not necessarily correct and sometimes obviously inadequate.

The two multivariate techniques most commonly encountered in biological research are "principal component analysis", described in detail by Holland (1969a), and "canonical analysis". A simple example will be given of each.

An example of principal component analysis

Several good examples of principal component analysis have been presented by Moore (1965), who has kindly made available the details of one example so that they can be explained here. In brief, he was concerned to study the effect of the size of apple trees at planting on their future development. He had 54 maiden trees of the variety Worcester Pearmain grafted on the rootstock M.VII, and four measures of the size of each tree at planting, namely:

W tree weight,
X square of the trunk circumference,
Y lengths of laterals, and
Z length of central leader.

Clearly he could have carried out his study in quadruplicate, but this would have been laborious. Also, it could have led to confusion; although in general the four variates are in good agreement, the occasional disagreement could have indicated something of prime importance for the future.

Accordingly Moore began his study by clarifying the relationships between his four variates. First he worked out the standard deviation of each and used it as the unit of measurement. In this way he overcame any problems of scaling arising from the diverse units originally used and produced four new variates which will here be written w, x, y and z. With these new variates, all variances equal one and all covariances equal the correlation coefficient of the two variates concerned. In fact **A**, the matrix of variation actually observed, equals

$$\begin{pmatrix} 1{\cdot}0000 & & & \\ 0{\cdot}7571 & 1{\cdot}0000 & & \\ 0{\cdot}7792 & 0{\cdot}6684 & 1{\cdot}0000 & \\ 0{\cdot}5500 & 0{\cdot}6017 & 0{\cdot}3027 & 1{\cdot}0000 \end{pmatrix}$$

Note that, for the rest of this paper, the upper triangle of symmetrical matrices will be suppressed; also, in this example the rows and columns always refer to w, x, y and z in that order.

Turning now to **B**, this may be taken as

$$\begin{pmatrix} 1 & & & \\ 0 & 1 & & \\ 0 & 0 & 1 & \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

That is to say, as now measured, each variate is assumed to be subject to about the same error of observation, thus giving rise to equal diagonal elements. Moreover the errors are taken to be independent, thus making all elements off the diagonal equal to zero. It may be objected that because the diagonal elements are all the same this does not mean that they all equal one, but a constant multiplier will be absorbed into the latent roots. Since only the ratios of latent roots are of interest, a constant multiplier will have no effect.

The characteristic of principal component analysis is the adoption of this value of **B** after adjusting the variates to make it reasonable. Adjustment may be made in various ways according to the nature of the problem. In essence, then, the enquiry concerns sources of variation other than those due to errors of observation.

The next step is to find the latent roots which a computer shows to be 2·855, 0·721, 0·253 and 0·170. Two things may be noted about them. One is that their sum, as always in component analysis, equals the trace of **A**; the other that component analysis ideally leads to a series of nearly equal latent roots, here 0·253 and 0·170, together with others that are inflated to various degrees by sources of variation present in **A** but ignored in **B**.

Here there are two such latent roots, of which the more important is 2·855 with a characteristic vector shown by the computer to be

$$(0{\cdot}55 \quad 0{\cdot}54 \quad 0{\cdot}49 \quad 0{\cdot}42)$$

Note that in component analysis the elements of a characteristic vector are conventionally scaled so that the sum of their squares equals one. Since only the ratio of the elements has been of importance so far, this does not affect any conclusions. As to the vector itself, it suggests that there are sources of variation that affect all variates to something like the same degree in terms of standard deviations. This need surprise nobody, but the derivation of the vector in precise terms provides a measure of size that integrates all the original ones, namely

$$V_1 = 0{\cdot}55w + 0{\cdot}54x + 0{\cdot}49y + 0{\cdot}42z$$

which could perhaps more conveniently be written in terms of W, X, Y and Z.

Incidentally, the variance of a quantity so obtained always equals the latent root, which here is 2·855 or 71·4 per cent of the total variation as measured by the trace of **A**.

The second inflated root is 0·721 or 18·0 per cent, which corresponds to the characteristic vector

$$(-0{\cdot}16 \quad 0{\cdot}06 \quad -0{\cdot}56 \quad 0{\cdot}81)$$

This suggests a variate, V_2, equal to

$$-0{\cdot}16w+0{\cdot}06x-0{\cdot}56y+0{\cdot}81z$$

whose variability shows that in some trees there must be a marked imbalance between length of laterals and length of leader, which again is hardly surprising, though it is useful to have a precise measure of the attribute.

Quite apart, however, from giving quantitative form to two tree characteristics, the existence of which could have been surmised, the analysis has shown that there is no need to seek further explanation of the observed variation. There remains only $0{\cdot}253+0{\cdot}170 = 0{\cdot}423$ or 10·6 per cent, which appears to be associated with uninflated latent roots and therefore due chiefly to experimental error. In fact, the quantities derived from the last two latent roots may be regarded as virtually constant, at least as far as this particular nursery is concerned. They are

$$V_3 = 0{\cdot}35w-0{\cdot}84x+0{\cdot}26y+0{\cdot}31z$$
$$V_4 = 0{\cdot}74w+0{\cdot}02x-0{\cdot}61y-0{\cdot}28z$$

with variances of 0·253 and 0·170 respectively. The constancy of V_3 shows that any tree with a high value for x must have had generally high values of w, y and z as well, while the constancy of V_4 shows a similar balance between w on the one hand and y and z on the other.

As a means of discerning natural relationships, these results are to be preferred to those obtained from regression techniques. If, however, the task were to estimate one variate on the basis of the others, regression methods would be correctly used.

It would be wrong to suggest that results always work out so neatly. Indeed, the ready intelligibility of V_3 and V_4 can only be due to good fortune. Whenever two or more invariances are established they can be combined to give others, all equivalent to the original and all equally valid. Thus, in this instance, if someone had argued from the virtual constancy over all trees of V_3 and V_4 that $5V_3+3V_4$, i.e.

$$3{\cdot}9w-4{\cdot}14x-0{\cdot}53y+0{\cdot}81z$$

must be nearly constant also, he would obviously be justified. Further, his assertion would have an intelligible meaning because it implies a balance between w and x. If now he proceeded to consider the quantity $3V_3-5V_4$, which is independent of the first, he would obtain

$$-2{\cdot}65w-2{\cdot}62x+3{\cdot}83y+2{\cdot}33z$$

which establishes the invariance of a balance between w and x on the one hand and y and z on the other. Both methods express the same results, but one may be preferred as giving a simpler way of visualizing them. The vectors derived from the computer are not necessarily the best, and sometimes appear to be complicated and arbitrary, yet consideration may suggest a simple means of presentation. Combining vectors may be applied to those that show the form of variation as well as those that show invariances.

An example of canonical analysis

Elsewhere (Pearce, 1965a) data have been presented giving

p = log(circumference of the base of the trunk)
q = log(circumference of the top of the trunk)
r = log(weight of the tree above ground)

for 36 plum trees of the variety Pershore Egg, which had just been grubbed at an age of 20 years. Nine had been grown on their own roots, and nine had been grafted on each of the rootstocks Common Plum, Common Mussel and Myrobalan B.

From the data two matrices of variances and covariances can be obtained. One of them,

$$\mathbf{A} = \begin{pmatrix} 0{\cdot}068008 & & \\ 0{\cdot}065494 & 0{\cdot}063214 & \\ 0{\cdot}137955 & 0{\cdot}132854 & 0{\cdot}280736 \end{pmatrix}$$

with three degrees of freedom, represents the variation between the mean values of the variates for the root stocks, i.e. the rootstock effects added to environmental differences. The other,

$$\mathbf{B} = \begin{pmatrix} 0{\cdot}001169 & & \\ 0{\cdot}001110 & 0{\cdot}001296 & \\ 0{\cdot}002475 & 0{\cdot}002515 & 0{\cdot}006508 \end{pmatrix}$$

with 32 degrees of freedom, represents the variation between trees having the same rootstock, i.e. variations due to environmental effects alone. This provides a characteristic example of canonical analysis. One matrix represents the effects of treatments as confused by experimental error; the other the error isolated.

The data were originally presented to illustrate component analysis by a study of **B**. Briefly, 95·0 per cent of the total variation was ascribed to a single vector, which showed that the three variates changed together, with displacements proportionate to their standard deviations. In fact, the environmental effects that operated in this experiment led only to differences in general size of the plants and not to any marked changes in form. If any further effects remained to be discovered, they could account for only 5·0 per cent of the total variation.

Turning to the effects of rootstocks, it may first be noted that they caused general size differences of the same kind as those that arose from environmental variation, though of much greater extent. This is shown by Table IIA, in which the values for rootstock means were obtained by deriving a variate from the first vector given by component analysis, evaluating it for each tree and working out an analysis of variance on the results.

The purpose of canonical analysis is however different, being to discern what there is in **A** that is absent in **B**, i.e. in this instance to find differences induced by rootstocks that could not have been brought about by environment. This is done by finding the three latent roots, λ, that makes $\mathbf{A} - \lambda\mathbf{B}$ singular, i.e. 58·82, 0·92 and 0·48. Clearly the first has been inflated by some effect, the characteristic vector being

$$(5{\cdot}59 \quad 0{\cdot}69 \quad -0{\cdot}52)$$

Again rootstock means are given in Table IIA. The vector represents individual behaviour on the part of the first variate, i.e. the circumference at the bottom of the trunk, and is quite foreign to the variation brought about by the environment, in which all variates change together.

Table IIA. *Plum rootstock experiment. Values of vectors obtained (a) from component analysis, representing general size, and (b) from canonical analysis, representing individual behaviour of the lower trunk girth*

Rootstock	(a)	(b)
Own roots	55·6	9·15
Common Plum	51·3	8·61
Common Mussel	55·3	9·07
Myrobalan B	60·1	9·71
Standard error	0·51	0·059
F-value	50·88	58·82

The investigation may be continued by examining the vectors associated with the other latent roots, namely

$$\begin{matrix} & (-1{\cdot}18 & -8{\cdot}40 & 4{\cdot}57) \\ \text{and} & (13{\cdot}01 & -8{\cdot}39 & -2{\cdot}43) \end{matrix}$$

Here it should be explained that, once the stage of interpretation is reached, there is little difference between component and canonical analysis. Consequently the vectors could be studied as in the previous example, where the object was to express the results in a way that should be biologically meaningful. Equally the method to be given here could be used with component analysis.

It follows from the vectors that

$$\begin{aligned} V_2 &= -1{\cdot}18p - 8{\cdot}40q + 4{\cdot}57r \\ \text{and} \quad V_3 &= 13{\cdot}01p - 8{\cdot}39q - 2{\cdot}43r \end{aligned}$$

are both nearly constant as between rootstocks. The situation can be simplified by eliminating q and asserting the invariance of $8{\cdot}39V_2 - 8{\cdot}40V_3$, which equals

$$-119{\cdot}2p + 58{\cdot}8r$$

Dividing by $-58{\cdot}8$ it appears that

$$2{\cdot}01 \log(\text{lower trunk circumference}) - \log(\text{tree weight})$$

is virtually the same for all rootstocks. Taking antilogarithms a rule appears, namely, that the weight of tree above ground is proportionate to (lower trunk circumference)$^{2{\cdot}01}$, i.e. according to a square law. (It would, however, be rash to base much on results from only four rootstocks.)

Again, a distinction of cases needs to be emphasized. If it were a question of estimating tree weight in a particular instance, it would be necessary to use regression methods so as to allow for error. Where, however, the object is to discern the natural law, obscured as it is by error, the correct technique is canonical analysis.

Before leaving the subject, certain differences from component analysis should be noted. Thus, the elements of a vector are not scaled so that their squares sum to one but according to a more complicated convention. Also, the sum of the latent roots is no longer necessarily equal to the trace of **A**, because in canonical analysis each latent root equals the F-value in the analysis of variance of the corresponding variate. This can be confirmed by examination of Table IIA. These are minor differences, however; the two techniques have much in common.

Conclusions

In scientific work no one is surprised when the interpretation of results is shown to depend partly upon the intuition of an individual. It is indeed recognized that some people have a knack, which others lack, of seeing to the heart of a problem. Nevertheless, when confronted with results that have a precise mathematical form there is a tendency to complain if interpretation is needed and to decry the conclusions as lacking in objectivity. This is absurd. Conclusions suggested by electronic computers are like any other; they need to be interpreted and the interpretation needs to be tested. Thus the suggestion has been made that environmental variation in an apple nursery can lead to changes in general size and can disturb the balance between the growth of the central leader and the laterals, but that is all; other relationships are unaffected. Anyone who is sufficiently interested can put the matter to the test by applying various environmental factors under controlled conditions to see if the quantities that appear to be invariant can in fact be varied and, if so, by which factors. Again,

a relationship is suggested between the weight of the aerial parts of a plum tree and the area of the trunk just above ground level. The assertion can be tested by trying out more rootstocks, choosing especially those likely to give trees of unusual habit. This is the classical method of science.

There is no automated road to success in the generation of hypotheses, and multivariate analyses provide no certainties. They do nevertheless help in thinking about a problem because they can give precise form both to the ways in which the organism can be changed and those in which it will resist change. Used to clarify interrelations between the parts of an organism, multivariate methods can be invaluable.

Some applications are to be found in the following papers: Holland, 1966, 1967a, 1968d, 1969a; Moore, 1965; Pearce, 1959, 1965c; and Pearce and Holland, 1960, 1961.

Appendix III

CALCULATING THE ANALYSIS OF VARIANCE

Generalities

The designs described in Chapters 2 and 3 all generate data of a sort that is commonly dealt with using the technique known as the "analysis of variance". That is to say:

(i) the total variation among the data is first worked out as a sum of squared deviations;
(ii) that part of the total associated with the blocking systems, i.e. blocks or rows and columns, is removed;
(iii) the remainder is "partitioned" between error and the effects of treatments.

It should be emphasized that no valid method can exist for carrying out stage (iii) unless randomization procedures have been adequate, i.e. in a block design, for each block the treatments assigned to it should have been allocated at random to the plots. In a row-and-column design a satisfactory array of treatments should be written down and then both rows and columns of the array permuted at random to give the final allocation of treatments.

Parallel with the analysis of the variation is the analysis of the degrees of freedom, which provide a measure of the potential for variation among the quantities being studied. Usually they number one fewer than the quantities themselves, e.g. between six treatments means there will be five degrees of freedom, as described in Section 11.

Once the analysis is complete it can be used in many ways depending upon the purpose for which the data have been collected. Some of the possibilities have been set out elsewhere (Pearce, 1965, Chapter 2).

In the past it was usual to present a special algorithm for each design, but nowadays more general methods are available. The one to be described here was first advanced by Kuiper (1952) and has subsequently been developed by Corsten (1958) and Worthington (1975). In its mathematical aspects it has been studied by Caliński (1971) and Pearce, Caliński and Marshall (1975). It has been chosen here for two reasons. One is its genuinely wide availability. It is important to check that a sufficient randomization procedure has been followed but otherwise it is not necessary to understand all the niceties of the design. The other is its simplicity when applied with standard designs. For example, when the experiment is in randomized blocks, the arithmetic needed is little more than what is needed by an algorithm specific to that design.

For block designs it is available generally, though difficulties can arise if there is a "disconnection", i.e. if some treatments occur only in certain blocks, the other treatments occurring only in the other blocks, as happens, for example, when an interaction is confounded. In that case there is no information to be gained about the contrast between the two sets of treatments because it depends entirely upon the block means and their effect was eliminated at stage (ii). Any complications are best dealt with by analysing each part of the disconnection separately and combining them at the end or else by using an algorithm specific to the circumstances. Where a main effect has been confounded, as in a split-plot design, the difficulties are sometimes better dealt with using the methods set out in Section 36. Since mishaps in a block design, e.g. the loss of plots or the misapplication of treatments, usually lead to a design that is as valid as the original, though it may well be less suitable for the purposes of the experiment, it is usually best to analyse the data as they stand without reference to what was intended.

For a row-and-column design some further considerations apply. For one thing (Pearce, 1975), it is possible for some information about a contrast to be lost among the rows and the rest among the columns, thus producing all the effects of a disconnection though none is apparent in either rows or columns separately. The situation may not be discovered until the information matrix is calculated, which is a reason for finding it early rather than later.

Also, the method requires that rows and columns should be orthogonal to one another and some mishaps, e.g. the loss of a plot, lead to the violation of this requirement, so techniques such as those for fitting missing values are needed to retain orthogonality. Some other mishaps, e.g. the application to a plot of the wrong treatment, merely produce an alternative design.

Table IIIA. *Total spread in inches of twelve strawberry plants per plot.* (*Treatments A, B, C and D are the application of different herbicides. Treatment O is a clean cultivation.*) *Plots are presented according to the field plan*

	Block I	*Block II*	*Block III*	*Block IV*
	C, 107	A, 136	B, 118	O, 173
	A, 166	O, 146	A, 117	C, 95
	D, 133	C, 104	O, 176	C, 109
	B, 166	B, 152	D, 132	A, 130
	O, 177	D, 119	B, 139	D, 103
	A, 163	O, 164	O, 186	O, 185
	O, 190	D, 132	C, 103	B, 147
Totals	1102	953	971	942

Grand total 3968

Block designs

The method will be illustrated using the data set out in Table IIIA. It will be seen that there are five plots for each herbicide and eight for the control. Also, there is considerable balance in the design, since each herbicide occurs once in three blocks and twice in the fourth, while the control occurs twice in each block. The design is, in fact, in supplemented balance as described in Section 32.

The first task is to find the variation that remains after allowing for blocks. As measured by the sum of squared deviations it equals

$$(107^2+136^2+\ldots+147^2)-(1102^2+953^2+971^2+942^2)/7 = 585\,178{\cdot}00-564\,688{\cdot}29 = 20\,489{\cdot}71$$

The first term was found by summing the squares of the data, the second by squaring each block total, dividing by the number of plots in the block and then adding the quotients. There are 27 degrees of freedom among 28 plots and 3 among the four block means, so 24 remain. It will be noted that the sum of squared deviations has been calculated to two more decimal places than were given by squaring the data themselves. Assuming that the data are expressed reasonably the rule is a good one.

Stages (i) and (ii) having thus been accomplished, there remains the more difficult task of apportioning the quantity, 20 489·71, between treatments and error. For a start it is necessary to find the treatment totals, which are

A, 712; B, 722; C, 518; D, 619; O, 1397

It is also convenient at this stage to write down the "incidence matrix", which has a row for each treatment and a column for each block. It shows the number of times each treatment occurs in each block, i.e.

	I	II	III	IV
A	2	1	1	1
B	1	1	2	1
C	1	1	1	2
D	1	2	1	1
O	2	2	2	2

The adjusted treatment totals, **Q**, are now formed by taking each actual treatment total in turn and subtracting from it the block means, multiplying each by the number of times the treatment concerned occurred in that block, i.e.

$$\begin{aligned} Q_A &= 712-[(2\times 1102)+953+971+942]/7 \\ &= 712-(1102+3968)/7 = -12{\cdot}286 \\ Q_B &= 722-(971+3968)/7 = 16{\cdot}429 \\ Q_C &= 518-(942+3968)/7 = -183{\cdot}429 \\ Q_D &= 619-(953+3968)/7 = -84{\cdot}000 \\ Q_O &= 1397-(2\times 3968)/7 = 263{\cdot}286 \end{aligned}$$

A check is provided by adding the adjusted treatment totals and making sure that their sum is zero. It will be noticed that they have been calculated to three decimal places more than the data themselves.

The main part of the algorithm is the "iteration" set out in Table IIIB. (By iteration a mathematician means the repetition of an arithmetical routine that takes an approximate solution and improves upon it. When further repetition has no effect, it may fairly be concluded that a reasonably exact solution has been obtained. The iteration is then said to have "converged".) The object is to find the true effects of treatments after allowing for any disturbances brought about by blocks. Thus, in this experiment Treatment A is overweighted in Block I. Accordingly, if Block I gives better growth than the rest, Treatment A will appear better than it really is relative to the others. The need is to disentangle the two effects. An initial approximation is given by the vector, $\mathbf{v}_1$, found by dividing each adjusted treatment total by the number of plots that received the treatment. Thus, for Treatment A the initial guess is $-12{\cdot}286/5 = -2{\cdot}457$ and so on.

Table IIIB. *Calculations arising from the analysis of data in Table IIIA*

					$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_3$	$\mathbf{v}_4$
2	1	1	1	(5)	−2·457	0·306	0·009	0·000
1	1	2	1	(5)	3·286	0·470	0·013	0·000
1	1	1	2	(5)	−36·686	−0·672	−0·019	−0·001
1	2	1	1	(5)	−16·800	−0·104	−0·003	0·000
2	2	2	2	(8)	32·911	0·000	0·000	0·000
(7)	(7)	(7)	(7)					
1·530	−0·519	2·350	−3·360	$\mathbf{u}_1$				
0·044	−0·015	0·067	−0·096	$\mathbf{u}_2$				
0·001	0·000	0·002	−0·003	$\mathbf{u}_3$				
0·000	0·000	0·000	0·000	$\mathbf{u}_4$				

To start the iteration, the incidence matrix is written on the top left-hand corner, as in Table IIIB. The vector of treatment replications and $\mathbf{v}_1$ are written to the right of it and the vector of block sizes is written below it.

Now $\mathbf{v}_1$ is "projected", i.e. for Block I each element of $\mathbf{v}_1$ is multiplied by the number of times the treatment occurs in Block I, the products are added and the sum divided by the number of plots in the block, which gives

$$[(2\times -2{\cdot}457)+(3{\cdot}286)+(-36{\cdot}686)+(-16{\cdot}800)+(2\times 32{\cdot}912)]/7 = 1{\cdot}530$$

and similarly for the other blocks. The resulting values, which constitute $\mathbf{u}_1$, are written across the bottom.

Now $\mathbf{v}_2$ is formed by the "dual projection" of $\mathbf{u}_1$. In this operation, for Treatment A each element of $\mathbf{u}_1$ is multiplied by the number of times Treatment A occurs in the block concerned and the sum of products is divided by the treatment replication, i.e.

$$[(2\times 1{\cdot}530)+(-0{\cdot}519)+(2{\cdot}350)+(-3{\cdot}360)]/5 = 0{\cdot}306$$

and so for the other treatments. These quantities constitute $\mathbf{v}_2$, which is written to the right of $\mathbf{v}_1$. It can be shown mathematically that $(\mathbf{v}_1+\mathbf{v}_2)$ gives a better approximation to the effects of treatments than $\mathbf{v}_1$ alone.

The iteration now continues. From $\mathbf{v}_2$ a projection forms $\mathbf{u}_2$, which in turn yields $\mathbf{v}_3$ by a dual projection. A further projection and dual projection give $\mathbf{v}_4$. However, $\mathbf{u}_4$ proves to contain only zero values, so the iteration can terminate at this point with the conclusion that the effects of treatments, known as the "treatment parameters", are estimated by the elements of $(\mathbf{v}_1+\mathbf{v}_2+\mathbf{v}_3+\mathbf{v}_4)$, i.e. by

A, −2·142; B, 3·769; C, −37·378; D, −16·907; O, 32·911

The parameters can be checked by multiplying each by the corresponding treatment replication and adding the products, which should sum to zero, i.e.

$$(5\times-2{\cdot}142)+(5\times3{\cdot}769)+(5\times-37{\cdot}378)+(5\times-16{\cdot}907)+(8\times32{\cdot}911)=-0{\cdot}002$$

The result is sufficiently near zero to be accounted for by rounding-off errors. If the test is failed, the check can be applied to $\mathbf{v}_1$, $\mathbf{v}_2$, etc. in turn to find the point at which the fault crept in. (Sometimes rounding-off errors accumulate to the point where a constant needs to be added or subtracted throughout to satisfy the test.) If now the general mean (3968/28 = 141·714) is added to each parameter, the result will be the adjusted treatment means, viz.

A, 139·6; B, 145·5; C, 104·3; D, 124·8; O, 174·6

If the design had been orthogonal, e.g. in randomized blocks, all elements of $\mathbf{u}_1$ would have been zero, i.e. $\mathbf{v}_1$ is enough without $\mathbf{v}_2$, $\mathbf{v}_3$, etc. The iteration is therefore unnecessary and work is much reduced.

With the parameters known, it becomes possible to complete the analysis of variance. Two methods are available. The simpler is to obtain the sum of squared deviations for treatments by multiplying each adjusted treatment total by the corresponding parameter and adding the products, i.e. treatment *s-s* equals

$$(-12{\cdot}286\times-2{\cdot}142)+(16{\cdot}429\times3{\cdot}769)+(-183{\cdot}429\times-37{\cdot}378)$$
$$+(-84{\cdot}000\times-16{\cdot}907)+(263{\cdot}286\times32{\cdot}911)=17\,029{\cdot}64$$

Hence the sum of squared deviations for error must equal what is left, i.e.

$$20\,489{\cdot}71-17\,029{\cdot}64=3460{\cdot}07$$

A more lengthy method but one that provides useful supplementary information is to proceed by estimating the block parameters as well. As an approximation it is good enough to estimate each as the difference between the block mean and the general mean, i.e. for Block I to estimate the difference of that block from the area as a whole as

$$1102/7-3968/28=15{\cdot}714$$

Similar calculations for the other blocks suggest:

I, 15·714; II, −5·571; III, −3·000; IV, −7·143

However, just as the treatment effects have been disturbed by differences between blocks so have block effects been disturbed by differences between treatments. Furthermore, just as $(\mathbf{v}_1+\mathbf{v}_2+\ldots)$ gives the adjustment for treatments so $(\mathbf{u}_1+\mathbf{u}_2+\ldots)$ gives the corresponding adjustment for blocks, albeit with a negative sign. Hence, the block parameters may be estimated as

I, 14·139; II, −5·037; III, −5·419; IV, −3·684

A similar check applies as for the treatment parameters. If each is multiplied by the corresponding block size and the products are added, the sum should come to zero apart from rounding-off errors, which should be small. The sum here is −0·007. If the test fails it can be applied to the original estimate of the block effects and to $\mathbf{u}_1$, $\mathbf{u}_2$, etc. separately. It appears that Block I gives more growth than the other three.

With both sets of parameters known, it is a simple matter to find the "residual" for each plot. Thus, for the first plot in Table IIIA, it comes from Block I and received Treatment C. Hence, it would be expected to give a value of

$$141{\cdot}714 + 14{\cdot}139 - 37{\cdot}378 = 118{\cdot}475$$

In fact it gave 107, a discrepancy (a "residual") of −11·475. Table IIIC sets out the corresponding figures for all plots. It will be seen that they sum to zero over all blocks and all treatments. Also, they provide a means of studying the experimental area and suggesting which plots, if any, appeared to be faulty. Thus, there is a patch of poor growth at the top end of Blocks II and III and one of good growth at the bottom end of Blocks III and IV. It is not suggested that data should be censored on this basis, but useful hints can be obtained for the future.

Table IIIC. *Residuals corresponding to the data in Table A*

Block I		*Block II*		*Block III*		*Block IV*	
C,	−11·475	A,	1·465	B,	−22·064	O,	2·059
A,	12·289	O,	−23·588	A,	−17,153	C,	−5·652
D,	−5·946	C,	4·701	O,	6·794	C,	8·348
B,	6·378	B,	11·554	D,	12·612	A,	−5·888
O,	−11·764	D,	−0·770	B,	−1·064	D,	−18·123
A,	9·289	O,	−5·588	O,	16·794	O,	14·059
O,	1·236	D,	12·230	C,	4·083	B,	5·201
Totals	0·007		0·004		0·002		0·004

Treatment totals

A, 0·002; B, 0·005; C, 0·005; D, 0·003; O, 0·002

The residuals have, however, a special importance because the sum of their squares equals the sum of squared deviations for error. (Indeed, the residuals are estimates of the deviations referred to.) Hence, the previous calculation is checked, 3460·26 being now ascribed to error and the remaining 17 029·45 to treatments. It will be seen that the number of decimal places used, large though it may have seemed, has not obviated rounding-off errors, though the agreement is good enough. The two methods between them, the one specifically calculating the treatment sum of squares and the other doing the same for the error, the total of the two being known, provide a powerful check on the accuracy of the arithmetic.

Since there are 4 degrees of freedom among the five treatment means and 24 for error and treatments together, there are 20 for error alone. Hence, the analysis of variance may be written thus:

Source	*d.f.*	*s.s.*	*m.s.*
Treatments	4	17 030	4257·5
Error	20	3 460	173·0
Within blocks	24	20 490	

It will be seen that the mean square for treatments is so much larger than that for error that they can scarcely be regarded as measures of the same quantity, though a formal investigation of the matter calls for an *F*-test. Also, the standard error, which is found as the square root of the error variance, equals 13·2 or about 9·3 per cent of the general mean, a figure that

is encouragingly low and gives confidence that little has gone wrong in the conduct of the experiment.

Standard errors, etc.

So far the calculations have provided the treatment means, adjusted as may be necessary for differences between blocks, and a measure of the experimental error. It remains only to find the standard error associated with each quantity, i.e. means or differences of means, that the experiment is intended to study. Here some further iterations are needed, but since the quantities sought relate to the design rather than to any particular body of data they can be carried out at the initiation of the experiment and the results kept until needed. Indeed, there are strong reasons for embarking upon them at an early stage.

The object of the further iterations is to build up column by column the "information matrix" of the design. It has a row and a column for each treatment and it indicates the precision with which each adjusted treatment mean is estimated and the degree of association between any pair of them.

To take Treatment A first, it has five plots. Accordingly in an orthogonal design its variance would have been the error variance divided by five and it would have had zero covariance with every other mean, i.e. it would have been uncorrelated with them. In the non-orthogonal design, however, the variance may be inflated by the adjustments, which are themselves estimated quantities. Also, the process of adjustment may lead to correlations. For example, if an adjustment is made to one mean, a contrary adjustment may be needed to another, thus inducing a negative correlation, or perhaps a similar adjustment will have to be made to another, thus inducing a positive one.

It happens (Caliński, 1971; Pearce, Caliński and Marshall, 1975) that the iteration used to adjust treatment means can also be used to adjust a column of the information matrix. There is, however, a minor difficulty. The values ($\frac{1}{5}$, 0, 0, 0, 0) fail the test given by the treatment replications, but that can be remedied by subtracting $\frac{1}{28} = 0{\cdot}0357$ from each (28 = the number of plots). The iteration then proceeds thus starting with $\mathbf{v}_1$ as ($\frac{1}{5} - \frac{1}{28}$, $-\frac{1}{28}$, $-\frac{1}{28}$, $-\frac{1}{28}$, $-\frac{1}{28}$):

	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_3$	*Correction*	*Column for A*
A	0·16429	0·00429	0·00012	0·03571	0·20441
B	−0·03571	−0·00143	−0·00004	0·03571	−0·00147
C	−0·03571	−0·00143	−0·00004	0·03571	−0·00147
D	−0·03571	−0·00143	−0·00004	0·03571	−0·00147
O	−0·03571	0·00000	0·00000	0·03571	0·00000

The subtracted constant passes through the iteration unchanged and can therefore be added on again at the end. Repeating the iteration for the other treatments gives the following as the information matrix:

	A	B	C	D	O
A	0·2044	−0·0015	−0·0015	−0·0015	0·0000
B	−0·0015	0·2044	−0·0015	−0·0015	0·0000
C	−0·0015	−0·0015	0·2044	−0·0015	0·0000
D	−0·0015	−0·0015	−0·0015	0·2044	0·0000
O	0·0000	0·0000	0·0000	0·0000	0·1250

Anyone who noticed that the four herbicide treatments were of equal status and that the control was orthogonal to blocks could have saved himself some trouble, but the inexperienced should beware of short cuts. The symmetry of the matrix provides a useful check in all cases.

It follows that the standard error of the adjusted mean for a treatment such as A, B, C or D is

$$\text{square root } (0{\cdot}2044 \times 173{\cdot}0) = 5{\cdot}95$$

and for the control, O, it is

$$\text{square root } (0{\cdot}1250 \times 173{\cdot}0) = 4{\cdot}65$$

The figure 173·0 is, of course, the error variance derived in the analysis of variance.

However, the standard errors of individual means do little to assist in this instance, because the experiment has not been carried out in standardized conditions. More interest attaches to differences between means, especially to the difference between a herbicide treatment and the control, where the standard error is

$$\text{square root } \{[0{\cdot}2044 + 0{\cdot}1250 - 2(0{\cdot}0000)] \times 173{\cdot}0\} = 7{\cdot}55$$

Multiplying by t for 20 degrees of freedom and a significance level of 0·05, i.e. $t = 2{\cdot}086$, it appears that a difference greater than 15·7 should be accounted significant at that level, so that it may reasonably be concluded that all the herbicides did harm to the plants, at least by reducing spread.

In this investigation the main object was not to compare the herbicides among themselves, but it will help to illustrate the use of the information matrix to note that the standard error for such a comparison would have been

$$\text{square root } \{[0{\cdot}2044 + 0{\cdot}2044 - 2(-0{\cdot}0015)]\, 173{\cdot}0\} = 8{\cdot}44$$

which gives a significant difference ($P = 0{\cdot}05$) of 17·6. It should, however, be used bearing in mind the problems raised by unnominated effects as set out in Sections 71 and 72.

Row-and-column designs

The method described above is readily adapted for use with row-and-column designs. Let there be b rows and c columns so that the total number of plots is bc. Also, let the block totals of data come to B_1, B_2, etc. up to B_b and let the column totals similarly be C_1, C_2, up to C_c. If, further, G is the grand total of data, the sum of squared deviations to be apportioned between error and treatments is

$$\text{Sum of (data)}^2 - (B_1^2 + B_2^2 + \ldots + B_b^2)/c - (C_1^2 + C_2^2 + \ldots C_c^2)/b + G^2/(bc)$$

with $(b-1)(c-1)$ degrees of freedom.

Otherwise everything is much as for a block design, though there are three differences that should be noted.

(i) The calculation of **Q** is rather more complicated. To take a simple example, suppose the design were as follows ($b = 2, c = 4$):

C	B	A	B
B	A	C	A

It is too small to be of much use, but it will serve for purposes of illustration. Taking T_A to be the total of data from Treatment A,

$$Q_A = T_A - (B_1 + 2B_2)/4 - (C_2 + C_3 + C_4)/2 + 3G/8$$

The coefficients of $B_1/4$ and $B_2/4$, the row means, represent the number of times the treatment occurs in each row and similarly those of $C_1/2$, $C_2/2$, $C_3/2$ and $C_4/2$, the column means, refer to the incidence of the treatment on to four columns. The coefficient of $G/8$, the general mean, is the treatment replication. Similarly,

$$Q_B = T_B - (2B_1 + B_2)/4 - (C_1 + C_2 + C_4)/2 + 3G/8$$

$$Q_C = T_C - (B_1 + B_2)/4 - (C_1 + C_3)/2 + 2G/8$$

Where a treatment occurs equally often in each row, as is the case with C, the expression for its Q will reduce to what it would have been if the only blocking system had been that of columns, i.e. $(B_1+B_2)/4 = 2G/8$. Similarly, if a treatment occurs equally often in columns, it is sufficient to consider only rows. If it occurs equally often in rows and also in columns, the expression for Q reduces to $T-rG/(bc)$, where T is the treatment total and r is its replication. (The same thing happened in a block design when a treatment was orthogonal to blocks.)

Table IIID. *Form of iteration for a row-and-column design*

Rows		*Columns*					$\mathbf{v}_1$	$\mathbf{v}_2$
1	2	0	1	1	1	(3)	0·20833	0·05556
2	1	1	1	0	1	(3)	−0·12500	−0·02778
1	1	1	0	1	0	(2)	−0·12500	−0·04167
(4)	(4)	(2)	(2)	(2)	(2)			
−0·04167	0·04167	−0·12500	0·04167	0·04167	0·04167	$\mathbf{u}_1$		

(ii) The iteration requires the use of two incidence matrices, that for treatments relative to rows and that for treatments relative to columns. They are written down side by side and treated as if they formed a single matrix. Thus, if it were a question of finding the variance and covariances of Treatment A, the iteration would begin as in Table IIID. It has not been continued, but anyone who takes it further will notice the slowness of the convergence. The reason is twofold. For one thing, there are two non-orthogonalities to be considered; for another, the experiment is very small. Nevertheless, if the iteration is carried far enough it will converge. In this instance, eight or nine cycles are needed and that in practice is about the upper limit. Usually three or four suffice.

The check on treatment values is the same as before. The values of the horizontal vectors should sum to zero for both rows and columns, as happens for $\mathbf{u}_1$.

If treatments are orthogonal to either rows or columns the incidence matrix for that classification can be omitted, since no adjustment on its account is called for more, nor will one be made if it is included. If treatments are orthogonal to both rows and columns, as in a Latin square, the iteration is not needed anyway, $\mathbf{v}_1$ being sufficient.

(iii) Once the iteration is complete, parameters for both rows and columns can be found as before by adding $\mathbf{u}_1$, $\mathbf{u}_2$, etc. and subtracting the sum from the unadjusted values obtained by taking the differences between each row (or column) mean and the general mean. (Of course, if no iteration has been necessary, no adjustment is needed.) There is, however, an important difference. In a block design the residual equals

datum − general mean − block parameter − treatment parameter

but in a row-and-column design it equals

datum − general mean − row parameter − column parameter − treatment parameter

Apart from these three differences, everything proceeds as for a block design.

Appendix IV

THE ANALYSIS OF COVARIANCE WITH A NOTE ON THE PARTITIONING OF THE TREATMENT VARIATION

In the analysis of covariance it is supposed that the quantity under study is subject to disturbance by some other quantity, which is distributed over the plots in a manner unrelated to treatments. Thus, in an analysis of the weight of fruit in an experiment it might be expected that the crop from a plot depended to some extent upon the size of the trees at the beginning of the experiment before the treatments were applied and it might seem desirable to make allowance for those initial differences.

In such cases the quantity under study is termed the "dependent variate" and will here be denoted by the symbol, y. The quantity causing the disturbance is termed the "independent variate" and will be written, x. The assumptions that must be justified before undertaking an analysis of covariance do not usually cause much trouble. They may be set out thus:

(i) The independent variate, x, is independent of the treatments. Ideally x should be measured before any treatments have been applied. If x is related to the treatments, it must be recognized that the elimination of its effect eliminates also part of the treatment effect (see Section 80).

(ii) The dependent variate, y, gives the same error for each treatment (see Section 73), or it must be transformed, if need be, to ensure that this is so. No similar assumption need be made with regard to x.

(iii) The relationship between x and y (or the transformed value of y) must be of the form, $y = a + bx$, where a and b are constant but unknown numbers. To make sure that this relationship holds it may be necessary to transform x, though small departures from the relationship are usually not serious. Even if the relationship has some other form, covariance will do no harm though it may not do any good.

(iv) The value of b must be independent of the treatments, i.e. if a change in x produces a certain effect on y for plots receiving a certain treatment, it must produce the same effect for any other treatment. If x or y or both have been transformed, this property should hold for the transformed variates, not the untransformed.

Before proceeding to the details of the analysis of covariance, it may be worth while to explain a little about the association of two variates in the simple case where there is no classification by blocks, rows, columns or treatments but only a single series of paired observations. For the moment it will not matter which is the dependent variate and which the independent. The following figures represent x, the minimum air temperature at East Malling for the week preceding the time of writing this appendix, and y, the corresponding minimum on grass, both in °F.

x,	42	51	56	55	50	50	59,	Total 363,	Mean 51·9
y,	40	48	54	52	48	48	55,	Total 345,	Mean 49·3

The first step is to work out the sum of data squared for x and to subtract a correction term, and similarly for y. This gives

$$C_{xx} = (42^2 + 51^2 + \ldots + 59^2) - \tfrac{1}{7}(363)^2 = 182{\cdot}86$$

$$C_{yy} = (40^2 + 48^2 + \ldots + 55^2) - \tfrac{1}{7}(345)^2 = 153{\cdot}43$$

C_{xy} is then worked out by using the product of x and y in every case instead of x^2 or y^2, thus:

$$C_{xy} = (42.40 + 51.48 + \ldots + 59.55) - \tfrac{1}{7}(363)(345) = +166{\cdot}29$$

The sign is added because C_{xy} can be positive or negative, unlike C_{xx} and C_{yy}, which are always positive.

It is now possible to work out the "correlation coefficient" between x and y. It measures the degree of association between the two and equals $C_{xy}/\sqrt{(C_{xx}.C_{yy})}$; so in the example it equals +0·9928. The value varies from −1 for complete negative association i.e. an increase in x means a decrease in y and vice versa, the value of one exactly determining the value of the other) by way of zero in the case of no association to +1 in the case of complete positive association (i.e. an increase in one means an increase in the other, the value of one exactly determining the value of the other). This particular value is high and shows that the two variates are associated positively but not quite completely (i.e. the value of one does not quite determine the value of the other, though it nearly does so).

Correlation deals with the closeness of the association between the two variates; it does not say what the relationship is, that being the province of "regression". The regression coefficient of y on x is $C_{xy}/C_{xx} = +0{\cdot}9094 = b$,* i.e. if x changes its value by one, y may be expected to change its value by b. Thus, in this example, on the first day, when x differed from its mean by (42−51·9) °F = −9·9 °F, it is to be expected that y will differ from its mean by (−9·9)(+0·9094) °F = −9·0 °F. That gives (49·3−9·0) °F as the grass minimum temperature to be expected on a day with an air minimum of 42 °F. Actually, the figure was 40 °F, which is as near as could be expected. The correlation coefficient, though high, was not equal to one, and that gives warning that the grass minimum cannot be calculated exactly from the air minimum, but other factors must operate as well.

Similarly, the regression of x on y is $C_{xy}/C_{yy} = 1{\cdot}0838$. Consequently two lines can be drawn on a graph, one showing the value of y to be expected for each x and the other the value of x to be expected for each y. For cases of incomplete association they will not be the same, and care is sometimes needed in choosing the right one in any particular instance.

The significance of correlation coefficients can be judged from tables (Fisher and Yates, 1938) or by the following analysis of variance. (r is the correlation coefficient, n is the number of pairs of observations upon which it is based.)

Source	*d.f.*	*s.s.*	*m.s.*	*F*
Correlation	1	r^2	r^2	$\frac{(n-2)r^2}{(1-r^2)}$
Error	$n-2$	$1-r^2$	$(1-r^2)/(n-2)$	
Total	$n-1$	1		

The method is based on an important result, namely, that the variation in y after allowance has been made for x is given by the sum of squares, $C_{yy}(1-r^2) = C_{yy} - C_{xy}^2/C_{xx}$, with one degree of freedom fewer than C_{yy}.

The analysis of covariance is a development of regression methods and, like them, has the advantage of making no assumptions as to the exact value of b, the change brought about in y by a unit change in x. Thus, it becomes possible to adjust y by x on the basis of the data itself. Given that a tree gave 200 kg of fruit over a four-year period of calibration, there is no obvious means of estimating how much fruit it would give over the six-year period of the subsequent experiment and any estimate that was made would be regarded with some suspicion by experienced investigators. If regression methods are used, the relationship derived is known to be appropriate to the circumstances. In particular, if it happens that for some reason a useless calibrating variate has been chosen, little harm will have been done, because in that case b comes out to be zero or something small. Accordingly, with the method of covariance

* Some people find the nomenclature a little confusing, the effect of x on y being measured by the regression of y on x.

it is permissible to "take a chance" on an independent variate that may have something to do with the problem though the experimenter cannot be certain. If it is irrelevant, the resulting adjustments will be negligible; if relevant, something will have been learnt about a further possibility in calibration.

It is now possible to return to the problem of the analysis of covariance. The data in Table IVA are taken from a cover-crop trial conducted at East Malling some years ago with the apple variety Barnack Beauty. The data are laid out in a column, because that was the design of the experiment in the field and it is helpful when residuals are to be evaluated to see how they were distributed over the experimental area. The variate for analysis, y, was the crop in pounds over the four-year period following the application of treatments. Adjustment was needed on x, the crop in bushels over the four years immediately preceding application.

Table IVA. *Data and residuals from a cover crop experiment on apples with four blocks and six treatments*

		Data		*Residuals*		
Block	*Treatment*	x	y	x	y	adj y
I	A	8·2	287	0·9583	42·750	15·533
	F	7·6	222	−0·4917	−17·250	−3·285
	C	6·8	234	−0·3417	−1·000	8·705
	B	8·2	271	1·0583	43·500	13·443
	D	5·7	189	−1·0167	−41·000	−12·125
	E	6·1	210	−0·1667	−27·000	−22·266
II	D	5·5	205	−1·6167	−38·000	7·916
	F	10·1	301	1·6083	48·750	3·073
	A	9·4	290	1·7583	32·750	−17·187
	E	7·0	276	0·3333	26·000	16·534
	B	6·0	209	−1·5417	−31·500	12·286
	C	7·0	210	−0·5417	−38·000	−22·615
III	B	9·1	243	−0·0084	−17·667	−17·428
	F	9·0	238	−1·0584	−34·417	−4·357
	D	10·2	312	1·5166	48·833	5·760
	C	9·7	286	0·5916	17·833	1·031
	A	7·7	254	−1·5084	−23·417	19·423
	E	8·7	279	0·4666	8·833	−4·419
IV	B	10·1	348	0·4917	5·667	−8·298
	A	8·5	307	−1·2083	−52·083	−17·766
	C	9·9	371	0·2917	21·167	12·882
	F	10·5	357	−0·0583	2·917	4·573
	E	8·1	344	−0·6333	−7·833	10·153
	D	10·3	375	1·1167	30·167	−1·548

Block totals	*I*	*II*	*III*	*IV*
x	42·6	45·0	54·4	57·4
y	1413	1491	1612	2102

The nature of the treatments was such that partitioning was called for, the five degrees of freedom between the six treatments having little collective meaning. In fact, Treatments A–E represented different species or mixtures of species sown to effect soil cover, while Treatment F represented clean cultivation. It is, therefore, better to ask two questions and not one, namely,

(i) Does it matter what is sown so long as the soil surface is covered with vegetation? For this purpose the relevant degrees of freedom are the four between Treatments A–E.

(ii) Supposing that it is possible to generalize about the effects of cover crops, do they differ in their effect from leaving the soil surface clean? The question here relates to the contrast between Treatment F and the rest, which has one degree of freedom.

It will be noted that Question (i) is logically anterior to (ii); if cover crops do differ appreciably among themselves it makes no sense to ask if they differ collectively from clean cultivation, an enquiry that should provoke the counter question, "Which cover crop have you in mind?"

The computations have several stages.

1. *Computation of the analysis of variance for each variate separately*

That is readily done. The sum of squared deviations to be partitioned between error and treatments is

$$8{\cdot}2^2+7{\cdot}6^2\ldots+10{\cdot}3^2-(42{\cdot}6^2+45{\cdot}0^2+54{\cdot}4^2+57{\cdot}4^2)/6 = 31{\cdot}6267 \quad \text{for } x$$

and

$$287^2+222^2+\ldots+375^2-(1413^2+1491^2+1612^2+2102^2)/6 = 24\,181{\cdot}67 \quad \text{for } y$$

The next step is to work out the values of **Q** for both variates. Since everything is orthogonal, for any treatment

$$Q = T-rm$$

where T is the data total for the treatment, r is the treatment replication, which is four in all cases, and m is the general mean of data (8·3083 for x and 275·750 for y). Also, the elements of $\mathbf{v}_1$, are those of **Q** divided by the treatment replication, so the following table is readily derived:

	x		y	
Treatment	**Q**	$\mathbf{v}_1$	**Q**	$\mathbf{v}_1$
A	0·5668	0·1417	35·000	8·750
B	0·1668	0·0417	−32·000	−8·000
C	0·1668	0·0417	−2·000	−0·500
D	−1·5332	−0·3833	−22·000	−5·500
E	−3·3332	−0·8333	6·000	1·500
F	3·9668	0·9917	15·000	3·750

Since everything is orthogonal there is no need to iterate, the elements of $\mathbf{v}_1$ being estimates of the treatment parameters without need of adjustment. The treatment sums of squares are now found by multiplying each element of **Q** by the corresponding parameter and adding the products. Thus, for x the sum of squared deviations comes to

$$(0{\cdot}5668\times 0{\cdot}1417)+(0{\cdot}1688\times 0{\cdot}0417)+\ldots(3{\cdot}9668\times 0{\cdot}9917) = 7{\cdot}3933$$

and for y it equals

$$(35{\cdot}000\times 8{\cdot}750)+(-32{\cdot}000\times -8{\cdot}000)+\ldots+(15{\cdot}000\times 3{\cdot}750) = 749{\cdot}50$$

At this point it would be possible to complete the analysis of variance, though it is usually better to carry out the check described in Appendix III by working out the residuals.

To do so, it is necessary to know the block parameters also. In an orthogonal design they are readily found, being equal in each instance to the difference between the block mean and the general mean. (If the design had been non-orthogonal the adjustments could have been derived from the iteration.)

The block parameters are accordingly estimated as

	I	II	III	IV
x	−1·2083	−0·8083	0·7583	1·2583
y	−40·250	−27·250	−7·0833	74·5833

Hence, the residuals given in Table IVA are found without difficulty. Thus, the first entry, that for Treatment A in Block I is $8{\cdot}2-8{\cdot}3083-(-1{\cdot}2083)-0{\cdot}1417$ for x and similarly for y. The sums of squared residuals are

$$x,\ 24{\cdot}2333;\quad y,\ 23\,432{\cdot}17$$

2. *Computation of the analysis of covariance*

At a number of points it has been necessary to square quantities relating to x and then to square the corresponding quantities of y. What is needed now is to multiply the two together. Thus, the sum of products to be partitioned is

$$(8{\cdot}2\times 287)+(7{\cdot}6\times 222)+\ldots+(10{\cdot}3\times 375)$$
$$-[(42{\cdot}6\times 1413)+(45{\cdot}0\times 1491)+(54{\cdot}4\times 1612)+(57{\cdot}4\times 2102)]/6=+710{\cdot}100$$

Unlike a sum of squares, a sum of products can be either positive or negative.

Then, to obtain the treatment sum of squares each element of **Q** was multiplied by the corresponding parameter, both relating to the same variate. Now, each element of **Q** for x is multiplied by the corresponding treatment parameter for y. Alternatively, and this provides a check, the elements of **Q** for y can be multiplied by the parameters for x, i.e. the treatment sum of products is

$$(0{\cdot}5668\times 8{\cdot}750)+(0{\cdot}1668\times -8{\cdot}000)+\ldots+(3{\cdot}9668\times 3{\cdot}750)=+21{\cdot}850$$

or

$$(35{\cdot}000\times 0{\cdot}1417)+(-32{\cdot}000\times 0{\cdot}0417)+\ldots+(15{\cdot}000\times 0{\cdot}9917)=+21{\cdot}850$$

Finally, the error sums of squares were found by squaring residuals, either for x or for y. What is now needed are the products of residuals, i.e. the error sum of products is

$$(0{\cdot}9583\times 42{\cdot}750)+(-0{\cdot}4917\times -17{\cdot}250)+\ldots+(1{\cdot}1167\times 30{\cdot}167)=+688{\cdot}250$$

Accordingly, the analyses of variance and covariance can be written down thus:

Source	*d.f.*	x^2	xy	y^2
Treatments	5	7·3933	+21·850	749·50
Error	15	24·2333	+688·250	23 432·17
Sum	20	31·6267	+710·100	24 181·67

Everything checks very well apart from a small rounding-off error in the analysis of variance for x. The figures in the error line following the degrees of freedom are respectively C_{xx}, C_{xy} and C_{yy}, so b, which equals C_{xy}/C_{xx}, is +28·4010. Hence, an additional bushel of apples from a plot during the calibrating period implies an additional 28·40 pounds of fruit during the period of the trial, a result that could not have been obtained in any other way.

With this figure known, various adjustments become possible. One is shown in the last column of Table IVA, where each residual for x has been multiplied by 28·4010 and subtracted from the corresponding residual for y. The result is a residual for y adjusted by the previous cropping performance of the plot and therefore made up from the variation that has come in since the calibration finished with the initiation of the experiment.

Similarly, the treatment means can be adjusted. Thus, to take Treatment A as an example, the mean value of x was 8·45, whereas the general mean was 8·31, so the treatment had been placed by the randomization on trees rather better than average. A better figure is therefore obtained by asking what should be expected of average trees, i.e. by subtracting $0{\cdot}14\times 28{\cdot}4$

(= 4·0) from the mean actually obtained, an adjustment that reduces 284·5 to 280·5. Similar adjustments give means as follows:

B, 266·6; C, 274·1; D, 281·0; E, 300·8; F, 251·4

It now appears that F was not really a very productive treatment, though it appeared to have been so on account of the good trees to which it had been assigned. It should be noted that it is not necessary to adopt average trees as the standard. With pseudo-variates especially, some conventional standard may be preferred. For example, a disease might break out during the course of an experiment and all the plots might be graded for symptom expression, 0 for no symptoms and 1, 2, 3, ... for symptoms of increasing seriousness. At the end the experimenter might prefer to adjust all means of the dependent variate, y, to a standard of apparently healthy plots, i.e. to x equals zero, rather than to a standard representing average symptom expression, i.e. to all x made equal to the mean value.

Another important adjustment is that of the analysis of variance of y, the dependent variate. After allowing for the effect of x, C_{yy} becomes $C_{yy}-C_{xy}^2/C_{xx}$ with one degree of freedom fewer. Hence in the error line 23 432·17 becomes 3885·18 with 14 degrees of freedom, while in the line for treatment and error together 24 181·67 becomes 8234·12 with 19 degrees of freedom. Hence the analysis of variance for y adjusted by x is

Source	*d.f.*	*s.s.*	*m.s.*
Treatments	5	4352·94	870·59
Error	14	3885·18	277·51
Sum	19	8238·12	

It will be noted that the treatments line is found by difference and in no other way. Also, the adjusted error sum of squares equals the sum of squares of the adjusted residuals of y, i.e. $15{\cdot}533^2+(-3{\cdot}285)^2+\ldots+(-1{\cdot}548)^2$.

It is incidentally permissible to use the quantities obtained to test whether the correlation between x and y is large enough to be accounted significant, the basis being the error line of y before and after adjustment, i.e.

Source	*d.f.*	*s.s.*	*m.s.*
Correlation	1	19 546·99	19 546·99
Error	14	3 885·18	277·51
Sum	15	23 432·17	

showing that the reduction of error brought about by adjustment is indeed significant. The test is the same as that given for the significance of a correlation coefficient. In this particular instance there was no call to use it, because it is part of the lore of those who work with apples that the two variates are in fact associated, but if the independent variate had been something of a shot in the dark, a digression might have been worth while to ensure that no further time should be wasted on the independent variate if it were in fact irrelevant.

A similar digression may be made to test whether the treatments differ in respect of x. The analysis of variance for x is worked out first and the mean square for treatments compared with that for error using the F-test. If the disparity is too large, for reasons given in Section 41 it may be unwise to adjust y by x. As a matter of fact, in the example of the cover-crop experiment, the mean of x for Treatment F does stand away from the rest and adjustment for the difference is needed if the effect of clean cultivation is to be rightly assessed. Here the position is clear because x was measured before the experiment began and cannot have been affected by the application of treatments. Also, the design was randomized so the values of x could not have affected the allocation of plots to treatments, but it is not difficult to visualize

situations in which an important difference depended in large measure upon calculated adjustments.

3. *Partitions and their validity*

In many instances, if not most, in which the analyses of variance and covariance are used there will be occasion to partition the treatment line. Only rarely will it be reasonable to regard all treatment comparisons as an amorphous group; indeed, it is a sign of a young statistician's developing powers when he becomes conscious of the structure of his set of treatments and wants to look at certain comparisons to the exclusion of others.

Such a situation does not always call for a formal partition. Thus, in Appendix III, where the important contrasts were those of the four herbicides with the untreated control, it was sufficient to work out a least significant difference and to use it where it was appropriate. In the case of a factorial design, on the other hand, it is almost invariable practice to isolate individual degrees of freedom or groups of degrees of freedom that add up to the total for treatments.

However, a partition of degrees of freedom does not necessarily correspond to a valid partition of the sum of squared deviations, because the comparisons to be studied may not be independent of one another. Such was the case, for example, in Appendix III, where all comparisons involved the untreated control, so any error of estimation of its mean would have affected all the comparisons under study similarly. The first question then concerns what partitions may or may not validly be made.

Here it may be mentioned that a quantity derived from the treatment means may conveniently be expressed by a vector (Sections 20 and 36). Thus, if there were four treatments, A, B, C and D, and if it were a question of comparing B with the mean of A and C, D not entering into the immediate enquiry, the vector would be

$$(-\tfrac{1}{2} \quad 1 \quad -\tfrac{1}{2} \quad 0)$$

Since only the ratios of the elements matter, it could be written equally as well with

$$\begin{array}{ll} & (\tfrac{1}{2} \quad -1 \quad \tfrac{1}{2} \quad 0) \\ \text{or} & (1 \quad -2 \quad 1 \quad 0) \end{array}$$

the last being the most convenient and therefore the most usual, but nothing depends upon the choice.

Where replication is unequal, finding a partition that is both valid and meaningful can be a severe test of skill, but where all treatments are equally replicated several standard situations exist. For example, one way is continually to divide the set of treatments into smaller and smaller groups. Given six of them, the five degrees of freedom could be allocated thus:

$$\begin{array}{rrrrrr} (1 & 1 & 1 & 1 & 1 & -5) \\ (1 & 1 & 1 & 1 & -4 & 0) \\ (1 & 1 & 1 & -3 & 0 & 0) \\ (1 & 1 & -2 & 0 & 0 & 0) \\ (1 & -1 & 0 & 0 & 0 & 0) \end{array}$$

That is to say, the last treatment is first compared with the first five as an undivided group, then the last but one with the first four and so on. There is no need always to separate one treatment at a time. The following is equally possible:

$$\begin{array}{rrrrrr} (1 & 1 & 1 & 1 & -2 & -2) \\ (1 & 1 & 1 & -3 & 0 & 0) \\ (1 & 1 & -2 & 0 & 0 & 0) \\ (1 & -1 & 0 & 0 & 0 & 0) \\ (0 & 0 & 0 & 0 & 1 & -1) \end{array}$$

Here a division was effected between the fourth and fifth treatment, the two groups thus formed being then dealt with as before. It will be noticed that the only quantities being studied are those that correspond to vectors with elements that sum to zero. Such quantities are known technically as "contrasts".

A more usual way of partitioning is to correspond to a factorial set of treatments. Suppose, for example, that the six treatments were not A–F but A1, B1, C1, A2, B2, C2, the five degrees of freedom would correspond to vectors, as follows:

Main effect of letters

$$(1 \quad 1 \quad -2 \quad 1 \quad 1 \quad -2)$$
$$(1 \quad -1 \quad 0 \quad 1 \quad -1 \quad 0)$$

Main effect of numbers

$$(1 \quad 1 \quad 1 \quad -1 \quad -1 \quad -1)$$

Interaction

$$(1 \quad 1 \quad -2 \quad -1 \quad -1 \quad 2)$$
$$(1 \quad -1 \quad 0 \quad -1 \quad 1 \quad 0)$$

It will be seen that the vectors for interaction contrasts were obtained by taking all combinations of a vector for one main effect and a vector for the other main effect and multiplying corresponding elements. That fits in, incidentally, with the rule that the number of degrees of freedom for an interaction is the product of the numbers of degrees of freedom for the two main effects.

It is all very well to write down sets of contrasts, but will their sums of squared deviations add up correctly? The answer depends upon the information matrix of the proposed design. To return to the apple cover-crop experiment, the design was very simple, being in four randomized blocks, so each treatment mean had a variance of (error mean square/4) and zero covariance with every other treatment mean, i.e. the information matrix is

$$\begin{pmatrix} \frac{1}{4} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{4} \end{pmatrix}$$

The partition required is first to consider the four degrees of freedom between treatments A–E and then the contrast of F with A–E, i.e. the scheme is the first of those given above, but with the first vector coming last and the others forming an undifferentiated group. To find out if two contrasts are independent it is necessary to write one down the side of the information matrix and the other across the top as a prelude to adding the triple products of elements of the matrix and the corresponding elements of the vectors, e.g.

$$\begin{array}{c} \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ -5 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 1 & 1 & 1 & -4 & 0 \end{array} \\ \begin{pmatrix} \frac{1}{4} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{4} \end{pmatrix} \end{array}$$

The sum of triple products is

$$(\tfrac{1}{4}\times 1\times 1)+(0\times 1\times 1)+\ldots+(0\times -5\times -4)+(\tfrac{1}{4}\times -5\times 0)=0$$

Because the sum comes to zero, the two contrasts will be estimated independently by the proposed design and so will all other pairs. Hence the partition can validly be made. Incidentally, so many of the triple products are themselves zero that a formal calculation of their sum is unnecessarily tiresome. To a practised eye the independence of the five contrasts is obvious.

4. *Computations needed to effect a partition*

A partition should be selected because it corresponds to the objectives of the experiment and for no other reason. When the most desirable partition has been discerned, a design should be adopted that will as far as possible enable it to be effected. Sometimes, as in the example in Appendix III, the important contrasts are inherently dependent, in which case no valid partition exists, but ideally a design should make possible a partition in which the sum of squared deviations for treatments breaks down into components, each of which is meaningful and corresponds to contrasts relevant to the purposes of the experiment.

In the present instance the only need is to isolate that part of the treatment variation that relates to the contrast between Treatment F and the rest. The method to be illustrated derives from the paper of Pearce, Caliński and Marshall (1974).

First a design constant is found for the contrast by summing triple products when the vector that expresses the contrast is written both vertically and horizontally, i.e.

$$\begin{array}{c} \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ -5 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 1 & 1 & 1 & 1 & -5 \end{array} \\ \begin{pmatrix} \frac{1}{4} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{4} \end{pmatrix} \end{array}$$

It is, in fact, $7\frac{1}{2}$. To consider now the contrast for x, it is evaluated by multiplying each treatment parameter by the corresponding element of the contrast vector and adding the products, i.e. it is

$$0{\cdot}1417+0{\cdot}0417+0{\cdot}0417-0{\cdot}3833-0{\cdot}8333+(-5)(0{\cdot}9917)=-5{\cdot}9500$$

and for y it is

$$8{\cdot}750-8{\cdot}000-0{\cdot}500-5{\cdot}500+1{\cdot}500+(-5)(3{\cdot}750)=-22{\cdot}500$$

Hence the contribution made by the contrast to the treatment sum of squares for x is $(-5{\cdot}9500)^2/7\frac{1}{2}$, i.e. 4·7203, and for y it is $(22{\cdot}500)^2/7\frac{1}{2}$, i.e. 67·50, while its contribution to the sum of products for xy is $(-5{\cdot}9500)(-22{\cdot}500)/7\frac{1}{2}$, i.e. +17·850, leading to the following partitioned analysis of variance and covariance:

Source	*d.f.*	x^2	xy	y^2
F versus A–E	1	4·7203	+17·850	67·50
Between A–E	4	2·6730	+4·000	682·00
Error	15	24·2333	+688·250	23 432·17
Sum	20	31·6267	+710·100	24 181·67

The line with four degrees of freedom for contrasts between A–E was found by difference and for most purposes that should suffice, but the suspicious may like to be satisfied that everything does in fact work out correctly. The calculations are not difficult.

Vector	Constant	Contrast in x	Contrast in y
(1 1 1 1 −4 0)	5	3·1750	−11·250
(1 1 1 −3 0 0)	3	1·3750	16·750
(1 1 −2 0 0 0)	1½	0·1000	1·750
(1 −1 0 0 0 0)	1½	0·1000	16·750

The sums of squares and products for the other four contrasts are therefore respectively

	x^2	xy	y^2
(1)	2·0161	−7·144	25·3125
(2)	0·6302	7·677	93·5208
(3)	0·0067	0·117	2·0417
(4)	0·0200	3·350	561·1250
	2·6730	4·000	682·0000

At this point it is possible to apply a covariance adjustment to the partitioned analysis of variance for y. Taking the differences between the five cover-crop treatments first, adding sums of squares and products into error gives

$$x^2,\ 26{\cdot}9063;\quad xy,\ 692{\cdot}250;\quad y^2,\ 24\ 114{\cdot}17$$

Each has 19 degrees of freedom. Adjusting in the usual way gives 6303·84 with 18 degrees of freedom. Subtracting the adjusted values for error alone (i.e. 3885·14 and 14) leaves a sum of squared deviations of 2418·70 with 4 degrees of freedom for differences between Treatments A–E, i.e. the mean-square is 2418·70/4 = 604·68. This is sufficiently larger than the error mean square ($F = 604{\cdot}68/277{\cdot}51 = 2{\cdot}18$) to raise doubts whether the cover crops really did have much the same effect, but the F-ratio is not significant even at the level, $P = 0{\cdot}05$, so it may well be decided that a generalization about cover crops is justified, which raises the further question whether cover crops differ in their effect from clean cultivation. Merging the line for the contrast of Treatment F with the others gives

$$x^2,\ 28{\cdot}9536;\quad xy,\ 706{\cdot}100;\quad y^2,\ 23\ 499{\cdot}67$$

with 16 degrees of freedom or 6279·80 with 15 degrees of freedom after adjustment, leaving 2394·62 with only 1 degree of freedom for the contrast. The F-ratio now equals 8·63 and there is indeed evidence of a real difference, the value being significant at $P = 0{\cdot}05$.

5. *Computation of least significant differences*

In this particular example significant differences are not required, the conclusions being clear without them. It is nevertheless necessary to explain how to calculate them, because this happy state of affairs does not always obtain.

The general expression for the difference between two adjusted means, one based on p plots and the other on q, with a difference of D in their mean values of x, is (Wishart, 1936)

$$t\sqrt{\left[\left(\frac{1}{p}+\frac{1}{q}+\frac{D^2}{C_{xx}}\right)\text{Error mean square}\right]}$$

where the error mean square is that in the analysis of y as adjusted by x.

Thus, in this example, the significant difference between the mean of A–E and that of F is calculated thus: The values of p and q are respectively 20 and 4. D equals 8·11–9·30 = −1·19 and C_{xx} has already been evaluated as 24·2334. The value of t for $P = 0{\cdot}05$ and 14 degrees of freedom is 2·145, while the error mean square is 277·52. That makes the significant difference 21·4. The actual difference exceeds this, being 280·7 − 251·4 = 29·3, but the

significance of the difference is not really in doubt, having already been demonstrated by the analysis of variance of y adjusted by x.

In order to avoid having to work out a fresh significant difference for each treatment comparison, it is usually convenient to work out the extreme significant differences between adjusted treatment means. Thus, in this example, if it had been necessary to work out the significance of all differences between the means of Treatments A–F it would have appeared that the smallest value of D was that between Treatments B and C, which give the same mean value for x. The largest value of D, on the other hand, is 1·82 between Treatments E and F. Consequently, the smallest significant difference is

$$t\sqrt{\left[\left(\frac{1}{4}+\frac{1}{4}+\frac{0^2}{24{\cdot}2334}\right)\text{Error mean square}\right]}=25{\cdot}3$$

and the largest is

$$t\sqrt{\left[\left(\frac{1}{4}+\frac{1}{4}+\frac{1{\cdot}82^2}{24{\cdot}2334}\right)\text{Error mean square}\right]}=28{\cdot}5$$

With this information, namely, that all the significant differences lie somewhere between 25·3 and 28·5, it is possible to proceed with little further computation. If there are any doubtful cases, the significant difference can be worked out for that particular value of D to clear up the obscurity.

This example well illustrates the twofold advantage of adjusting by the analysis of covariance in cases where there is a good degree of correlation between the two variates:

(1) It has led to a substantial reduction of error variation. The error mean square for y before adjustment was 1562·14, but after adjustment it was 277·52, so the precision of the trial had been increased to more than five times its former value (1562·14/277·52 = 5·63), the loss of one degree of freedom from the error being negligible in comparison.

(2) The adjustments to the treatment means have made clear the relative performances of the treatments despite an unlucky randomization that had assigned a disproportionate number of the good trees to the control.

Nevertheless, the analysis of covariance is not a panacea for all ills, its efficacy depending upon a suitable choice of independent variate.

Double covariance

It may well happen that an analysis of variance is further complicated by there being two independent variates. Thus, with strawberries it is not unusual to adjust crop after applying treatments, by both height and spread at the time of application. Again, it may be necessary to adjust both by a calibrating variate and pseudo-variate.

Although a double covariance adjustment is more laborious than a single one, it is not more difficult in conception. The method of computing results is very similar and the successive stages the same. In the following explanation the dependent variate will be called y and the two independent variates, w and x. Apart from there being three data per plot instead of two, leading to three sets of sums of squares and three of sums of products, everything is the same until Stage 4, the construction of the analysis or variance of y adjusted by w and x. With a double covariance the adjusted error sum of squares equals

$$C_{yy}-\frac{C_{ww}C_{xy}{}^2+C_{xx}C_{wy}{}^2-2C_{wx}C_{wy}C_{xy}}{C_{ww}C_{xx}-C_{wx}{}^2}$$

with *two* degrees of freedom fewer than those of C_{yy}. Similar expressions apply when error and treatments are merged.

At Stage 2 it is necessary to know the regression coefficients of y on w and x. The "partial regression coefficient of y on w", x remaining unaltered despite its association with w, is

$$\frac{C_{xx}C_{wy}-C_{wx}C_{xy}}{C_{ww}C_{xx}-C_{wx}^2}=b_w$$

Equally, the "partial regression of y on x", w remaining unaltered, is

$$\frac{C_{ww}C_{xy}-C_{wx}C_{wy}}{C_{ww}C_{xx}-C_{wx}^2}=b_x$$

With these quantities known it is possible to adjust on to the standard values of w and x. Thus, if w needs to be adjusted by d_w and x by d_x (care is needed here with signs), the adjusted value of y is

$$(\text{unadjusted mean of } y)-b_w d_w-b_x d_x$$

the adjustment having taken place in two steps.

At the end, in Stage 5, a further modification is needed, this time to work out the significant difference between two treatment means after adjustment, the one being based on p plots and the other on q. If the treatments differ by D_w in respect of w and by D_x in respect of x, the expression is (Wishart, 1936)

$$t\sqrt{\left[\left(\frac{1}{p}+\frac{1}{q}+\frac{D_w^2C_{xx}-2D_wD_xC_{wx}+D_x^2C_{ww}}{C_{ww}C_{xx}-C_{wx}^2}\right)\text{Error mean square}\right]}$$

where the error mean square is that in the analysis of variance of y adjusted by w and x.

It will have been seen from this that the computation of an analysis of variance with a double covariance adjustment calls for a lot of work, but is quite practicable if the effort is necessary.

REFERENCES

ABEYWARDENA, V. (1962a) Studies on biennial bearing tendency in coconut. *Ceylon Coconut Q.*, **13**, 112–25.

ABEYWARDENA, V. (1962b) Some suggestions for minimizing errors from mishaps in field experiments with perennials. *Ceylon Coconut Q.*, **13**, 126–34.

ABEYWARDENA, V. (1964) Statistical control of variability in coconut experiments. *Emp. J. exp. Agric.*, **32**, 166–74.

ABRAHAM, J. K. (1960) On an alternative method of computing Tukey's statistic for the Latin square model. *Biometrics*, **16**, 686–91.

ABRAHAM, T. P., HOOBAKHT, A. and NEJAD, P. A. (1971) Statistical studies on experimental techniques with the apple crop in Iran. *Exp. Agric.*, **7**, 249–55.

ABRAHAM, T. P., KHOSLA, R. K. and AGRAWAL, K. N. (1969) Uniformity trials on black-pepper, *Piper nigrum* L. *Indian J. agric. Sci.*, **39**, 780–806.

AGRAWAL, K. C. (1961a) Experimental designs at the Horticultural Research Institute, Saharanpur. *A.R. hort. Res. Inst., Saharanpur, 1960*, 100–9.

AGRAWAL, K. C. (1961b) On the determination of number of replications in fruit-tree experiments after taking into account the loss of trees in the course of experimentation. *Hort. Advance*, **5**, 230–3.

AGRAWAL, K. N., BAVAPPA, K. V. A. and KHOSLA, R. K. (1968) Study of size and shape of plots and blocks, and optimum number of pre-experimental periods on arecanut. *Indian J. agric. Sci.*, **38**, 444–60.

ALLEN, F. E. and WISHART, J. (1930) A method of estimating the yield of a missing plot in field experimental work. *J. agric. Sci.*, **20**, 399–406.

AMOS, J. and HOBLYN, T. N. (1928) Manurial trials with hops. *A.R. E. Malling Res. Stat. 1926–27. II Suppl.*, 165–71.

ANON. (1949) Notes and news. *Biometrics*, **5**, 256–7.

ANON. (1958) Study on uniformity data on pepper. *Statist. News Letter* (*Indian Council agric. Res.*) **8**, 6–7.

ANON. (1961) Experimentation on trees and other perennial crops. *Statist. News Letter* (*Indian Council agric. Res.*) **11**, 2–6.

ANON. (1962) Uniformity trial data on mango. *Statist. News Letter* (*Indian Council agric. Res.*) **11**, 11–5.

ANSCOMBE, F. J. (1948) The transformation of Poisson, binomial and negative binomial data. *Biometrika*, **35**, 246–54.

ASPIN, A. A. (1948) An examination and further development of a formula occurring in the problem of comparing two mean values. *Biometrika*, **35**, 88–96.

ASPIN, A. A. (1949) Tables for use in comparisons whose accuracy involves two variances, separately estimated. *Biometrika*, **36**, 290–3.

AWARTRAMANI, N. A. (1965) Glimpses of the way out of the impasse in the conduct of field experiments on coffee at Central Coffee Research Institute. *Indian Coffee*, **29**, 25–9, 34.

BAILEY, N. T. J. (1959) *Statistical methods in biology*. English Universities Press, London.

BAKER, G. A. and BAKER, R. E. (1953) Strawberry uniformity yield trials. *Biometrics*, **9**, 412–21.

BAKER, R. E. and BAKER, G. A. (1950) Experimental design for studying resistance of strawberry varieties to verticillium wilt. *Phytopathology*, **40**, 477–82.

BARTLETT, M. S. (1936) The square root transformation in analysis of variance. *J. roy. statist. Soc., Suppl.* **3**, 68–78.

BARTLETT, M. S. (1937) Some examples of statistical methods of research in agriculture and applied biology. *J. roy. statist. Soc., Suppl.* **4**, 137–83.

BARTLETT, M. S. (1938) The approximate recovery of information from replicated field experiments with large blocks. *J. agric. Sci.*, **28**, 418–27.

BARTLETT, M. S. (1947) The use of transformations. *Biometrics*, **3**, 39–52.

BARTLEY, B. G. D. (1970) Yield variation in the early productive years in trials with cacao (*Theobroma cacao* L.). *Euphytica*, **19**, 199–206.

BATCHELOR, L. D. and REED, H. S. (1918) Relation of the variability of yields of fruit trees to the accuracy of field trials. *J. agric. Res.*, **12**, 245–85.

BATEN, W. D. (1939) Formulas for finding estimates for two or three missing plots in randomized block layouts. *Tech. Bull. Mich. agric. Exp. Stat.* **165**.

BATEN, W. D. (1952) Variances of differences between means when there are two missing values in randomized block designs. *Biometrics*, **8**, 42–50.

BEAKBANE, A. B. (1967) A relationship between leaf structure and growth potential in apple. *Ann. appl. Biol.*, **60**, 67–76.

BEAKBANE, A. B. (1969) A new series of potential plum rootstocks. *A.R. E. Malling Res. Stat. 1968*, 81–3.

BERRY, G. (1967) A mathematical model relating plant yield with arrangement for regularly spaced crops. *Biometrics*, **23**, 505–15.

BERRY, G., CLEAVER, T. J., NELDER, J. A. and SALTER, P. J. (1966) Methods of recording data in the laboratory and field. *Exp. Agric.*, **2**, 69–80.

Bessis, R. (1960a) Deux méthodes rapides d'appréciation du nombre de fleurs dans les grappes de la vigne. *C. R. Acad. Agric. France*, **46**, 823–8.

Bessis, R. (1960b) Sur différentes modes d'expression quantitative de la fertilité chez la vigne. *C. R. Acad. Agric. France*, **46**, 828–32.

Blackith, R. E. (1957) Polymorphism in some Australian locusts and grasshoppers. *Biometrics*, **13**, 183–96.

Blackith, R. E., Davies, R. G. and Moy, E. A. (1963) A biometric analysis of development in *Dysdercus fasciatus* Sign. *Growth*, **27**, 317–34.

Blake, M. A. and Hervey, G. W. (1928) A standard for estimating the twig growth of one-year-old peach trees. *Bull. N.J. agric. Exp. Stat.* **475**.

Bleasdale, J. K. A. (1967a) The relationship between the weight of a plant part and total weight as affected by plant density. *J. hort. Sci.*, **42**, 51–8.

Bleasdale, J. K. A. (1967b) Systematic designs for spacing experiments. *Exp. Agric.*, **3**, 73–86.

Bleasdale, J. K. A. and Nelder, J. A. (1960) Plant population and crop yield. *Nature (Lond.)*, **188**, 342.

Bliss, C. I. (1937) The analysis of field experimental data expressed in percentages. (Russian.) *Pl. Prot., Leningrad*, **12**, 67–77.

Bliss, C. I. (1938) The transformation of percentages for use in the analysis of variance. *Ohio J. Sci.*, **38**, 9–12.

Bliss, C. I. (1967) *Statistics in biology, Volume I.* McGraw-Hill Book Co., New York.

Bliss, C. I. (1970) *Statistics in biology, Volume II.* McGraw-Hill Book Co., New York.

Bocquet, M. (1953) Note sur l'expérience de densité Marchal appliqué à l'hévéaculture. *Arch. Rubbercult.*, May 1953, Extra No., 194–9.

Borgman, H. H. (1953) Bepaling van de groei van vruchtbomen. *Meded. Dir. Tuinb.*, **16**, 239–42.

Bose, R. C., Shrikhande, S. S. and Parker, E. T. (1960) Orthogonal Latin squares and Euler's conjecture. *Canad. J. Math.*, **12**, 189–203.

Bose, S. S. (1938) The estimation of mixed-up yields and their standard errors. *Sankhyā*, **4**, 112–20.

Bose, S. S. and Mahalanobis, P. C. (1938) On estimating individual yields in the case of mixed-up yields of two or more plots in field experiment. *Sankhyā*, **4**, 103–11.

Boubals, D. (1957) L'organisation des vignes d'expérimentation. *Progr. agric. vitic.*, **148**, 31–49, 67–73.

Box, G. E. P. and Cox, D. R. (1964) An analysis of transformations. *J. roy. statist. Soc.*, **B26**, 211–52.

Bradford, F. C. (1944) Second-year changes in apparent vigor of apple varieties of prospective value as trunk-formers. *Proc. Amer. Soc. hort. Sci.*, **44**, 215–22.

Bradley, R. A., Walpole, R. E. and Kramer, C. Y. (1960) Intra- and inter-block analysis for factorials in incomplete block designs. *Biometrics*, **16**, 566–81.

Brase, M. A. and Tukey, H. B. (1936) The relation between size of apple seedling rootstocks and size of orchard tree. *Proc. Amer. Soc. hort. Sci.*, **34**, 298–304.

Brian, C. (1966) Contribution à l'interprétation des essais de porte-greffes du poirier. *Ann. Amél. Plantes*, **16**, 349–65.

Brunskill, R. T. (1957) A variable dosage sprayer for agricultural use. *J. agric. Engng Res.*, **2**, 135–40.

Butijn, J. and van der Boon, J. (1957) Correcties in opbrengsten van vruchtbomen in verband met de boomgroote. *Meded. Dir. Tuinb.*, **20**, 40–5.

Butters, R. (1964) Some practical considerations in the conduct of field trials with robusta coffee. *J. hort. Sci.*, **39**, 24–33.

Caliński, T. (1971) On some desirable patterns in block designs. *Biometrika*, **27**, 275–92.

Campbell, R. C. (1967) *Statistics for biologists.* Cambridge University Press.

Cannon, H. B. (1941) Studies in the variation of nursery fruit trees on vegetatively raised rootstocks. *J. Pomol. hort. Sci.*, **19**, 2–33.

Chandler, W. H. (1921) The trend of research in pomology. *Proc. Amer. Soc. hort. Sci.*, **18**, 233–40.

Chapas, L. C. (1961) Plot size and reduction of variability in oil-palm experiments. *Emp. J. exp. Agric.*, **29**, 212–24.

Chaturvedi, M. D. (1926) Measurements of the cubical contents of forest trees. *Oxford For. Mem.* **4**.

Cheesman, E. E. and Pound, F. J. (1932) Uniformity trials with cacao. *Trop. Agric., Trin.*, **9**, 277–88.

Christensen, J. R. (1948) Determinación de parcelas experimentales para viñas. *Experimenta Mendoza*, **1**, 20–5.

Clark, A. and Leonard, W. H. (1939) The analysis of variance with special reference to data expressed as percentages. *J. Amer. Soc. Agron.*, **31**, 56–66.

Clarke, G. M. (1967) Some problems in the use of calibrating variates. *A.R. Long Ashton agric. hort. Res. Stat. 1966*, 277–9.

Clarke, G. M. and Corke, A. T. K. (1955) Sources of variability in sampling blackcurrants for leaf spot. *A.R. Long Ashton agric. hort. Res. Stat. 1954*, 199–201.

Clarke, G. M. and Corke, A. T. K. (1956) A rapid method of assessing black currant leaf spot. *A.R. Long Ashton agric. hort. Res. Stat. 1955*, 196–200.

Clarke, R. T. (1963) A statistical design for experiments on airborne pests and diseases. *Trop. Agric., Trin.*, **40**, 67–9.

CLARKE, R. T. (1965) The presentation of numerical results of experiments for publication in a scientific agricultural journal. *Exp. Agric.*, **1**, 315–19.
COCHRAN, W. G. (1937) Problems arising in the analysis of a series of similar experiments. *J. roy. statist. Soc., Suppl.* **4**, 102–18.
COCHRAN, W. G. (1938) Some difficulties in the statistical analysis of replicated experiments. *Emp. J. exp. Agric.*, **6**, 157–75.
COCHRAN, W. G. (1951) Testing a linear relation among variances. *Biometrics*, **7**, 17–32.
COCHRAN, W. G. (1954) The combination of estimates from different experiments. *Biometrics*, **10**, 101–29.
COCHRAN, W. G. (1957) Analysis of covariance, its nature and uses. *Biometrics*, **13**, 261–81.
COCHRAN, W. G. and COX, G. M. (1950) *Experimental designs*. John Wiley & Sons, Inc., New York.
COCHRAN, W. G. and WATSON, D. J. (1936) An experiment on observer's bias in the selection of shoot heights. *Emp. J. exp. Agric.*, **4**, 69–76.
COLLISON, R. C. and HARLAN, J. D. (1930) Variability and size relations in apple trees. *Bull. N.Y. St. agric. Exp. Stat.* **164**.
COOKE, D. (1962) The timing of harvests in horticultural experiments. *Biometrics*, **18**, 478–98.
COONS, I. (1957) The analysis of covariance as a missing plot technique. *Biometrics*, **13**, 387–405.
CORNELL, J. (1974) More on extended block designs. *Biometrics*, **30**, 179–86.
CORNISH, E. A. (1940a) The estimation of missing values in incomplete block experiments. *Ann. Eugen.*, **10**, 112–18.
CORNISH, E. A. (1940b) The estimation of missing values in quasi-factorial designs. *Ann. Eugen.*, **10**, 137–43.
CORNISH, E. A. (1940c) The analysis of quasi-factorial designs with incomplete data. *J. Aust. Inst. agric. Sci.*, **6**, 31–9.
CORNISH, E. A. (1941) The analysis of quasi-factorial designs with incomplete data. 2. Lattice squares. *J. Aust. Inst. agric. Sci.*, **7**, 19–26.
CORSTEN, L. C. A. (1958) Vectors, a tool in statistical regression theory. *Meded. Landbouwhogeschool Wageningen*, **58**, 1–92.
COVAS, G. and CHRISTENSEN, J. R. (1945) Determinación del tamaño de parcelas para ensayos comparativos de rendimentos en la vid. *Rev. Argent. Agron.*, **12**, 26–9.
COX, C. P. (1968) Some observations on the teaching of statistical consultancy. *Biometrics*, **24**, 789–802.
COX, D. R. (1958) *Planning of experiments*. John Wiley & Sons, Inc., New York.
CROWTHER, E. M. (1947) Fertilizer experiment in colonial agriculture. *Colonial*, **214** (H.M.S.O., London), 10–28.
CUNNINGHAM, R. K. and BURRIDGE, J. C. (1959) The effect of the yield variability of amelonado cacao on the design of fertilizer field experiments in Ghana. *J. hort. Sci.*, **34**, 229–37.
CURNOW, R. N. (1961) Optimal programmes for varietal selection (with discussion). *J. roy. statist. Soc.*, **B23**, 282–318.
DAGNELIE, P. (1964) La détermination du nombre de répétitions en vue de l'estimation d'une moyenne. *Biométrie-Praximétrie*, **5**, 117–35.
DANCER, J. (1963) A note on the statistical analysis of yield differences in biennially bearing crops. *E. Afr. agric. For. J.*, **28**, 126–8.
DAS, M. N. and BHARGAVA, P. N. (1958) On missing and mixed-up plot techniques. *J. Ind. Soc. agric. Statist.*, **10**, 78–82.
DAS, M. N. and KULKARNI, G. A. (1958) On the accuracy of current approximate variances of treatment differences in randomized block designs with missing observations. *J. Ind. Soc. agric. Statist.*, **10**, 83–9.
DEDOLPH, R. R. (1960) A suggested method for handling data obtained with an exponential (variable dosage) sprayer. *Proc. Amer. Soc. hort. Sci.*, **75**, 789–98.
DEDOLPH, R. R., BASHAM, C. W. and STARK, F. C. (1960) An exponential sprayer for experimental work. *Proc. Amer. Soc. hort. Sci.*, **75**, 785–8.
DORSEY, M. J. and HOUGH, L. F. (1943) Relation between seedling vigor and tree vigor in apple hybrids. *Proc. Amer. Soc. hort. Sci.*, **43**, 106–14.
DUNCAN, D. B. (1955) Multiple range and multiple *F* tests. *Biometrics*, **11**, 1–42.
DUTTA, S. K. and HEATH, E. D. (1960) Size, shape and number of plots for field experiments with tea. *Emp. J. exp. Agric.*, **28**, 121–32.
DYKE, G. V. (1974) *Comparative experiments with field crops*. Butterworth, Sevenoaks.
DYSON, W. G. and FREEMAN, G. H. (1968) Seed orchard designs for sites with a constant prevailing wind. *Silvae Genetica*, **17**, 12–15.
EDEN, T. (1931) Studies in the yield of tea. I. The experimental errors of field experiments with tea. *J. agric. Sci.*, **21**, 547–73.
EDEN, T. (1935) Studies in the yield of tea. Pt III. Field experiments with potash and nitrogen in relation to the pruning cycle. *Emp. J. exp. Agric.*, **3**, 105–18.
EDEN, T. (1941) The selection of high yielding tea bushes for vegetative propagation. *Tea Q.*, **14**, 98–102.
EDGAR, J. L. (1938) Strawberry cultivation studies. II. Variability in individual plant size and cropping, with special reference to area and shape of plots for field experiments. *J. Pomol.*, **16**, 91–100.

ENGLEDOW, F. L. and YULE, G. W. (1926) The principles and practice of yield trials. *Emp. Cotton Growing Rev.*, **3**, 112–46.

ESQUIVEL, O. and SORIA, V. J. (1967) Estudio preliminar sobre el tamaño mínimo de parcela de cacao. *Proc. Carib. Reg. Amer. Soc. hort. Sci.*, **10**, 148–53.

EULER, L. (1782) Recherches sur une nouvelle espèce de quarrés magiques. *Verh. uitg. Zeeuwsch Genoot. Wet. Vlissingen*, **9**, 85–239.

FARAZDAGHI, H. and HARRIS, P. M. (1968) Plant competition and crop yield. *Nature* (*Lond.*), **217**, 289–90.

FEDERER, W. T. (1955) *Experimental design: theory and application.* Macmillan, New York.

FEDERER, W. T. (1961) Experimental error rates. *Proc. Amer. Soc. hort. Sci.*, **78**, 605–15.

FEDERER, W. T. and SCHLOTTFELDT, C. S. (1954) The use of covariance to control gradients in experiments. *Biometrics*, **10**, 282–9.

FERGUSON, J. H. A. (1962) Nomogram voor het onderscheidingsvermogen van de *F*-toets gemodificeerd naar Keuls. *Meded. Inst. Vered. TuinbGew., Wageningen*, **187** (also *Statist. Neerl.*, **16**, 177–80).

FERGUSON, J. H. A. (1963) Random variability in horticultural experiments. *Meded. Inst. Vered. TuinbGew., Wageningen*, **196** (also *Euphytica*, **11**, 213–20).

FINNEY, D. J. (1945a) Some orthogonal properties of the 4×4 and 6×6 Latin squares. *Ann. Eugen.*, **12**, 213–19.

FINNEY, D. J. (1945b) The fractional replication of factorial experiments. *Ann. Eugen.*, **12**, 291–301.

FINNEY, D. J. (1946a) Orthogonal partitions of the 5×5 Latin squares. *Ann. Eugen.*, **13**, 1–3.

FINNEY, D. J. (1946b) Orthogonal partitions of the 6×6 Latin squares. *Ann. Eugen.*, **13**, 184–96.

FINNEY, D. J. (1946c) Latin squares of the sixth order. *Experientia*, **2**, 404–5.

FINNEY, D. J. (1946d) Recent developments in the design of field experiments. III. Fractional replication. *J. agric. Sci.*, **36**, 184–91.

FINNEY, D. J. (1947) Statistical science and agricultural research. *Math. Gaz.*, **31**, 21–30.

FINNEY, D. J. (1948) Main effects and interactions. *J. Amer. statist. Ass.*, **43**, 566–71.

FINNEY, D. J. (1950) The fractional replication of factorial experiments—a correction. *Ann. Eugen.*, **15**, 276.

FINNEY, D. J. (1953a) *Statistical science in agriculture.* Oliver and Boyd, Edinburgh.

FINNEY, D. J. (1953b) Response curves and the planning of experiments. *Indian J. agric. Sci.*, **23**, 167–86.

FINNEY, D. J. (1955) *Experimental design and its statistical basis.* Cambridge University Press.

FINNEY, D. J. (1956a) The statistician and the planning of field experiments (with discussion). *J. roy. statist. Soc.*, **A119**, 1–27.

FINNEY, D. J. (1956b) Multivariate analysis and agricultural experiments. *Biometrics*, **12**, 67–71.

FINNEY, D. J. (1958a) Statistical problems of plant selection. *Bull. int. statist. Inst.*, **36**, 242–68.

FINNEY, D. J. (1958b) Plant selection for yield improvement. *Euphytica*, **7**, 83–106.

FINNEY, D. J. (1962) *An introduction to statistical science in agriculture.* Oliver and Boyd, Edinburgh.

FINNEY, D. J. (1964) The replication of variety trials. *Biometrics*, **20**, 1–15.

FINNEY, D. J. (1968) Teaching biometry in the university. *Biometrics*, **24**, 1–12.

FISHER, R. A. (1934) *Statistical methods for research workers.* Oliver and Boyd, Edinburgh.

FISHER, R. A. (1935) *The design of experiments.* Oliver and Boyd, Edinburgh.

FISHER, R. A. (1941) The negative binomial distribution. *Ann. Eugen.*, **11**, 182–7.

FISHER, R. A. (1948) Biometry. *Biometrics*, **4**, 217–19.

FISHER, R. A. (1954) The analysis of variance with various binomial transformations (with discussion). *Biometrics*, **10**, 130–51.

FISHER, R. A. and YATES, F. (1934) The 6×6 Latin squares. *Proc. Camb. phil. Soc.*, **30**, 492–507.

FISHER, R. A. and YATES, F. (1938) *Statistical tables for biological, agricultural and medical research.* Oliver and Boyd, Edinburgh.

FRAGA, C. G., JR., and CONAGIN, A. (1956) Delineamentos e analises de experimentos com cafeeiros. *Bragantia*, **15**, 177–91.

FREEMAN, G. H. (1953) Spread of diseases in a rectangular plantation with vacancies. *Biometrika*, **40**, 287–96.

FREEMAN, G. H. (1956) The selection and use of a panel for taste sensitivity tests with fruit. *A.R. E. Malling Res. Stat. 1955*, 86–8.

FREEMAN, G. H. (1959) The use of the same experimental material for more than one set of treatments. *Appl. Statist.*, **8**, 13–20.

FREEMAN, G. H. (1963) The combined effect of environmental and plant variation. *Biometrics*, **19**, 273–7.

FREEMAN, G. H. (1964) The use of a systematic design for a spacing trial with a tropical tree crop. *Biometrics*, **20**, 200–3.

FREEMAN, G. H. (1965) Methods of raising cocoa seedlings in the nursery and their effect on subsequent growth in the field. *J. hort. Sci.*, **40**, 341–9.

FREEMAN, G. H. (1967a) Problems in plant pathology and entomology trials. *Exp. Agric.*, **3**, 351–8.

FREEMAN, G. H. (1967b) The use of cyclic balanced incomplete block designs for directional seed orchards. *Biometrics*, **23**, 761–78.

FREEMAN, G. H. (1969) The use of cyclic balanced incomplete block designs for non-directional seed orchards. *Biometrics*, **25**, 561–71.

FREEMAN, G. H. (1973) Experimental designs with unequal concurrences for estimating direct and remote effects of treatments. *Biometrika*, **60**, 559–63.

FREEMAN, G. H. and THOMPSON, J. A. (1960) The assessment of wood growth on black currant bushes. *A.R. E. Malling Res. Stat. 1959*, 67–9.

FREEMAN, M. H. and TUKEY, J. W. (1949) Transformations related to the angular and the square-root *Mem. Rep. Princeton Univ. Statist. Res. Group*, **24**.

FREYMAN, S. and DOLMAN, D. (1971) A simple systematic design of planting density experiments with set row widths. *Canad. J. Plant Sci.*, **51**, 340–2.

GARNER, R. J. (1947) *The grafter's handbook*. Faber and Faber, London.

GARNER, R. J. and WALKER, W. F. (1938) The frameworking of fruit trees. *Occ. Pap. imp. Bur. Fruit Prod., E. Malling*, **5**.

GARRETSEN, F. and KRONENBERG, H. G. (1957) Opbrengstproeven met zwarte bessen. *Meded. Dir. Tuinb.*, **20**, 424–32.

GEISLER, G. and STAAB, J. (1958) Versuchsanstellung im Weinbau. *Vitis*, **1**, 257–81.

GILBERT, P. and GROSSMANN, B. (1961) Comparaison de deux méthodes d'analyse de données nonorthogonales dans le cas d'expériences à deux facteurs. *Biométrie-Praximétrie*, **2**, 19–36.

GILBERT, S. M. (1938a) Planning field experiments on *Coffea arabica*. *Trop. Agric., Trin.*, **15**, 16–18.

GILBERT, S. M. (1938b) Plot size in field experiments with *Coffea arabica*. *Trop. Agric., Trin.*, **15**, 52–5.

GLADWIN, F. E. (1924) Variation in plot yields in field experiments with the grape and their interpretation. *Proc. Amer. Soc. hort. Sci.*, **21**, 45–51.

GOODALL, D. W. (1954) Objective methods for the classification of vegetation. III. An essay on the use of factor analysis. *Aust. J. Bot.*, **2**, 304–24.

GOODCHILD, N. A. (1971) Designs having fruit trees as blocks and incorporating analysis of covariance. *Biometrics*, **27**, 313–23.

GOODCHILD, N. A. and SPRENT, P. (1966) The control of variability in raspberry experiments. *A.R. E. Malling Res. Stat., 1965*, 107–9.

GOODE, J. E. and MARCHANT, W. T. B. (1962) Use of mechanical barriers instead of guard rows in field experiments. *Nature (Lond.)*, **195**, 410.

GOULDEN, C. H. (1952) *Methods of statistical analysis*. John Wiley & Sons, Inc., New York.

GRÁCIO, A. M. (1969) Talhões experimentais para ensaios varietais em viticultura. Resultados de dois ensaios de uniformidade realizadoes em 1961. *Vinea Vino Portug. Docum. Sér. I*, **4**.

GREENHAM, D. W. P. (1965) A long-term manurial trial on dessert apple trees. *J. hort. Sci.*, **40**, 213–35.

GROSSE, B. (1941) Die Korrelation zwischen dem Fruchtertrag und dem vegetativen Wachstum bei Äpfeln. *Forschungsdienst*, **11**, 368–84.

HAINES, W. B. (1958) Réapparition cyclique de rendements élevés chez des palmiers analysés individuellement. *Oléagineux*, **13**, 367–73.

HAINES, W. B. (1959) The significance of cyclic peak yields in Nigerian oil palms. *Emp. J. exp. Agric.*, **27**, 1–9.

HAINES, W. B., CROWTHER, E. M. and THORNTON, G. J. (1954) Manuring hevea. V: Some long-term effects in the Dunlop (Malaya) experiments. *Emp. J. exp. Agric.*, **22**, 203–10.

HÄNNINGEN, P. (1951) Koeympyröiden käytöstä kenttäkokeissa. *Valt. Maatalousk. Tied*, **223**.

HATHEWAY, W. H. and WILLIAMS, E. J. (1958) Efficient estimation of the relationship between plot size and the variability of crop yields. *Biometrics*, **14**, 207–22.

HATTEMER, H. H. (1966) Die Eignung einiger Blatt- und Verzweigungsmerkmale für die Unterscheidung von Schwarzapfel Hybridklonen. *Züchter*, **36**, 317–27.

HATTON, R. G. (1931) The elimination of sources of error in field experiments. The standardization of fruit tree rootstocks. *A.R. E. Malling Res. Stat. 1928–30, II Suppl.*, 13–21.

HATTON, R. G., GRUBB, N. H. and KNIGHT, R. C. (1925) Black currant variety trials. Reliability of results. The variability in cropping of individual black currant bushes as a guide to the suitable size of experimental plots. *J. Pomol.*, **4**, 200–20.

HEALY, M. J. R. and WESTMACOTT, M. H. (1956) Missing values in experiments analysed on automatic computers. *Appl. Statist.*, **5**, 203–6.

HEDRICK, U. P. and ANTHONY, R. D. (1919) Twenty years of fertilizers in an apple orchard. *Bull. N.Y. agric. Exp. Stat.* **460**.

HEGEDIÜS, Á. (1969) The optimum plot size for field experiments in vineyards with high trained vines. *Szölö-Gyümölesterm*, **51**, 191–9.

HOBLYN, T. N. (1931) Field experiments in horticulture. *Tech. Commun. imp. Bur. Fruit Prod., E. Malling*, **2**.

HOBLYN, T. N. (1945) The design of field trials with cocoa. *Rep. Proc. Cocoa Res. Conf., London*, 164–8.

HOBLYN, T. N. (1958) Practical aspects of field experimentation with fruit. *A.R. E. Malling Res. Stat. 1957*, 43–7.

HOBLYN, T. N., GRUBB, N. H., PAINTER, A. C. and WATES, B. L. (1936) Studies in biennial bearing. I. *J. Pomol. hort. Sci.*, **14**, 39–76.

HOBLYN, T. N., PEARCE, S. C. and FREEMAN, G. H. (1954) Some considerations in the design of successive experiments in fruit plantations. *Biometrics*, **10**, 503–15.

HOFFMAN, M. B. (1930) The use of performance records in laying out a raspberry fertilizer experiment. *Proc. Amer. Soc. hort. Sci.*, **26**, 203–7.

HOLLAND, D. A. (1956) Soil acidity in relation to trunk girth of apples. *A.R. E. Malling Res. Stat. 1955*, 80–5.

HOLLAND, D. A. (1959) Studies in the measurement of apple trees. IV. The use of branch girths to estimate tree size. *A.R. E. Malling Res. Stat. 1958*, 77–9.

HOLLAND, D. A. (1966) The interpretation of leaf analyses. *J. hort. Sci.*, **41**, 311–29.

HOLLAND, D. A. (1967a) The interpretation of the chemical composition of some tropical crops by the method of component analysis. *Oléagineux*, **22**, 307–10.

HOLLAND, D. A. (1967b) Environmental modifications resulting from the culture and grubbing of trees. *J. appl. Ecol.*, **4**, 353–62.

HOLLAND, D. A. (1967c) Induced patterns of soil variation. *A.R. E. Malling Res. Stat. 1966*, 122–4.

HOLLAND, D. A. (1968a) Variation in chemical composition of apple leaves. *J. Sci. Fd Agric.*, **19**, 136–40.

HOLLAND, D. A. (1968b) Variation in chemical composition of orchard soils. *J. Sci. Fd Agric.*, **19**, 141–6.

HOLLAND, D. A. (1968c) The estimation of total leaf area on a tree. *A.R. E. Malling Res. Stat. 1967*, 101–4.

HOLLAND, D. A. (1968d) The component analysis approach to the interpretation of plant analysis data from groundnuts and sugar cane. *Exp. Agric.*, **4**, 179–85.

HOLLAND, D. A. (1969a) Component analysis—an approach to the interpretation of soil data. *J. Sci. Fd Agric.*, **20**, 26–31.

HOLLAND, D. A. (1969b) Variability in the yields of blackcurrants. *A.R. E. Malling Res. Stat. 1968*, 97–102.

HOLLAND, D. A. (1969c) Component analysis: an aid to the interpretation of data. *Exp. Agric.*, **5**, 151–64.

HOLLAND, D. A. (1971) Variability in the yields of strawberries. *J. hort. Sci.*, **46**, 55–62.

HOLLAND, D. A. and GREENHAM, D. W. P. (1958) The minimization of variability in experiments on apples planted on the site of a previous orchard. I. The effect of cultural treatments associated with three years' arable cropping. *J. hort. Sci.*, **33**, 153–70.

HOLLAND, D. A. and GREENHAM, D. W. P. (1962) The minimization of variability in apples planted on the site of a previous orchard. II. The effect of cultural treatments associated with seven years' arable cropping. *J. hort. Sci.*, **37**, 24–34.

HOLLAND, D. A. and GREENHAM, D. W. P. (1966) The minimization of variability in experiments on apples planted on the site of a previous orchard. III. A final appraisal of the effects of cultural treatments associated with three and seven years' arable cropping. *J. hort. Sci.*, **41**, 115–36.

HOLLAND, D. A., LITTLE, R. C., ALLEN, M. and DERMOTT, W. (1967) Soil and leaf sampling in apple orchards. *J. hort. Sci.*, **42**, 403–17.

HOLLIDAY, R. (1960) Plant population and crop yield. *Nature (Lond.)*, **186**, 22–4.

HOPP, H. (1954) A guide to extensive testing on farms. *U.S. Dep. Agric. foreign agric. Surv.*

HORSFALL, J. G. and RICH, S. (1949) Spirally arranged plots in a design for field assay of fungicides. *Bull. Conn. agric. Exp. Stat.* **530**.

HOYLE, M. H. (1973) Transformations—an introduction and a bibliography. *Int. statist. Rev.*, **41**, 203–23.

HUET, J. (1961) Les particularités de l'expérimentation fruitière. *Pomol. franç*, **3**, 225–9.

IRWIN, J. O. (1931) Precision records in horticulture. *J. Pomol. hort. Sci.*, **9**, 149–94.

IYER, G. C. (1968) Influences of plot sizes on precision in manuring experiments on *Hevea brasiliensis*. *J. Rubb. Res. Inst. Malaya*, **20**, 161–72.

IYER, T. A. G. (1958) Statistical analysis of experimental yield data from coconut trees. *Indian Coconut J.*, **11**, 106–24.

JEFFERS, J. R. N. and BLACK, T. M. (1963) An analysis of variability in *Pinus contorta*. *Forestry*, **36**, 199–218.

JOACHIM, A. W. R. (1935) A uniformity trial with coconuts. *Trop. Agriculturist*, **85**, 198–207.

JOHN, P. W. M. (1962) Testing two treatments when there are three experimental units in each block. *Appl. Statist.*, **11**, 164–9.

JOHN, P. W. M. (1963) Extended complete block designs. *Aust. J. Statist.*, **5**, 147–52.

JOLLY, A. L. (1942) Uniformity trials on estate cacao fields in Grenada, B.W.I. *Trop. Agric., Trin.*, **19**, 167–74.

JOLLY, G. M. and HOLLAND, D. A. (1958) Sampling methods for the measurement of extension growth of apple trees. *A.R. E. Malling Res. Stat. 1957*, 87–90.

JONES, W. W., EMBLETON, T. W. and CREE, C. B. (1957) Number of replications and plot sizes required for reliable evaluation of nutritional studies and yield relationships with citrus and avocado. *Proc. Amer. Soc. hort. Sci.*, **69**, 209–16.

KELLER, K. R. (1949) Uniformity trials on hops, *Humulus lupulus* L. *Agron. J.*, **41**, 389–92.

KENDALL, M. G. (1948) *The advanced theory of statistics*. Charles Griffin and Co., London.

KEULS, M. (1952) The use of the "Studentized range" in connection with an analysis of variance. *Euphytica*, **1**, 112–22.

KEULS, M. (1960) La puissance du critère *F* dans l'analyse de la variance de plans en blocs au hasard. Nomogrammes pour le choix du nombre de répétitions. *Biométrie-Praximétrie*, **1**, 65–80.

KOCH, E. J. and HUNTER, J. H. (1957) The use of experimental design and covariance techniques for increasing precision of pecan experiments. *Proc. Amer. Soc. hort. Sci.*, **69**, 170–5.

KOCH, E. J. and RIGNEY, J. A. (1951) A method of estimating optimum plot size from experimental data. *Agron. J.*, **43**, 17–21.

KRAMER, C. Y. and BRADLEY, R. A. (1957) Examples of intra-block analysis for factorials in group divisible, partially balanced, incomplete block designs. *Biometrics*, **13**, 197–224.

KUIPER, N. H. (1952) Variantie-analyse. *Statistica*, **6**, 149–94.

LAYCOCK, D. H. (1955) The effect of plot shape in reducing the errors of tea experiments. *Trop. Agric., Trin.*, **32**, 107–14.

LI, J. C. R. and KELLER, K. R. (1951) An application of serial correlation in field experiments. *Agron. J.*, **43**, 201–3.

LLEWELYN, F. W. M. (1968) The log(-log) transformation in the analysis of fruit retention records. *Biometrics*, **24**, 627–38.

LONGWORTH, J. F. and FREEMAN, G. H. (1963) The use of trunk girth as a calibrating variate for field experiments on cocoa trees. *J. hort. Sci.*, **38**, 61–7.

LOTODÉ, R. (1971) Possibilités d'amélioration de l'expérimentation sur cacaoyers. *Café, Cacao, Thé*, **15**, 91–104.

LOTT, W. L., SATCHELL, D. F. and HALL, N. S. (1950) A tracer-element technique in the study of root extension. *Proc. Amer. Soc. hort. Sci.*, **55**, 27–34.

LOVE, H. T., COWGILL, W. H. and HOPP, H. (1952) Test of a proposed method of sampling cinchona bark for fluorometric analysis for quinine. *Proc. Amer. Soc. hort. Sci.*, **60**, 279–82.

LUCAS, H. L. (1951) Bias in estimation of error in change-over trials with dairy cattle. *J. agric. Sci.*, **41**, 146–8.

LUTZ, H. (1939) The effect of size of young pecan trees on their subsequent growth and yield. *Proc. Amer. Soc. hort. Sci.*, **36**, 335–8.

MA, R. H. and HARRINGTON, J. B. (1949) A study of field experiments of semi-Latin square design. *Sci. Agric.*, **29**, 241–51.

MAGISTED, O. C. and FARDEN, C. A. (1934) Experimental error in field experiments with pineapples. *J. Amer. Soc. Agron.*, **26**, 631–44.

MAHLSTEDE, J. P. and LANA, E. D. (1958) Evaluation of the rooting response of cuttings by the method of ranks. *Proc. Amer. Soc. hort. Sci.*, **71**, 585–90.

MARCHANT, W. T. B. and BOA, W. (1962) An implement for laying plastic foil root barriers. *J. agric. Engng Res.*, **7**, 258–9.

MARTICOU, H. and MULLER, R. (1964) Essai de mise au point d'une méthode d'expérimentation adaptée aux conditions de la cacoyère camerounaise traditionnelle. *Café, Cacao, Thé*, **8**, 173–202.

MARTINEZ-ZAPORTA, F. (1952) Una contribución al empleo de los métodos estadísticos en fruticultura. *Bol. Inst. Invest. agron. Madrid*, **12**, 397–527.

MATÉRN, B. (1956) On the geometry of the cross-section of a stem. *Medd. SkogsforsknInst. Stockh.*, **46**, 5–28.

MATÉRN, B. (1970) Performance of various designs of field experiments when applied in random fields. (*Proc.*) *Third Conf. advis. Group Forest Statist., Jouy-en-Josas*, I.N.R.A., Paris, 119–28.

MAYER, R. *et al.* Méthodologie de l'expérimentation viticole. *Bull. Off. int. Vin.* **31** (336), 22–40; (337), 23–56; (338), 3–20.

MCHATTON, T. H. (1947a) The comparison of plot size in a peach experiment. *Proc. Amer. Soc. hort. Sci.*, **49**, 18–20.

MCHATTON, T. H. (1947b) Variation in peach yields. *Proc. Amer. Soc. hort. Sci.*, **49**, 121–4.

MCKENZIE, D. W. (1961) Rootstock–scion interaction in apples with special reference to root anatomy. *J. hort. Sci.*, **36**, 40–7.

MEAD, R. (1966) A relationship between individual plant-spacing and yield. *Ann. Bot.*, **30**, 301–9.

MEAD, R. (1967) A mathematical model for the estimation of interplant competition. *Biometrics*, **23**, 189–205.

MILLIKEN, G. A. and GRAYBILL, F. A. (1971) Tests for interaction in the two-way model with missing data. *Biometrics*, **27**, 1079–83.

MITSCHERLICH, E. A. (1934) Die Verarbeitung landwirtschaftlicher und anderer biologischer Versuchsergebnisse. *Schr. Königsberger Gelehrten Ges.*, **7**, 199–229.

MONROE, R. J. and MASON, D. D. (1955) Problems of experimental inference with special reference to multiple location experiments and experiments with perennial crops. *Proc. Amer. Soc. hort. Sci.*, **66**, 410–14.

MONZÓN, P. D., HADDAD, G. O. and FIGUEROA, R. M. (1969) Efectos de bordura y competencia en parcelas experimentales de parchita maracuyá (*Passiflora edulis* forma *flavicarpa* Degener.). *Agron. trop., Maracay*, **19**, 99–105.

MOORE, C. S. (1962) A comparison of the growth and cropping of apple trees of different habit. *A.R. E. Malling Res. Stat. 1961*, 87–90.

Moore, C. S. (1965) Inter-relations of growth and cropping in apple trees studied by the method of analysis. *J. hort. Sci.*, **40**, 133–49.

Moore, C. S. (1966) The relationship of increment in trunk girth with shoot growth in apple trees. *A.R. E. Malling Res. Stat. 1965*, 115–19.

Moore, C. S. (1968a) The course of variation in fruit tree experiments. *J. hort. Sci.*, **43**, 121–31.

Moore, C. S. (1968b) Response of growth and cropping of apples to treatments studied by the method of component analysis. *J. hort. Sci.*, **43**, 249–62.

Moore, C. S. (1975) Relative importance of rootstock and scion in determining growth and fruiting in young apple trees. *Ann. Bot.*, **39**, 113–23.

Moore, P. G. (1957) Transformations to normality using fractional powers of the variable. *J. Amer. statist. Ass.*, **52**, 237–46.

Morales, A. (1951) Análise estatística do ensaio de variedades de café: Parte II. *Bragantia*, **11**, 44–9.

Murray, R. K. S. (1934) The value of a uniformity trial in a field experimentation with rubber. *J. agric. Sci.*, **24**, 177–84.

Nair, K. R. (1940) The application of the technique of analysis of covariance to field experiments with several missing or mixed-up plots. *Sankhyā*, **4**, 581–8.

Narayanan, R. (1965) Some considerations in deciding the optimum number of recorded trees in a plot for experiments on immature *Hevea brasiliensis*. *J. Rubb. Res. Inst. Malaya*, **19**, 120–31.

Narayanan, R. (1966) Value of covariance analysis in manurial experiments on rubber. *J. Rubb. Res. Inst. Malaya*, **19**, 176–88.

Narayanan, R. (1968) Girth as a calibrating variate for improving field experiments on *Hevea brasiliensis*. *J. Rubb. Res. Inst. Malaya*, **20**, 130–5.

Narayanan, R. (1970) Double covariance analysis in manurial experiments on hevea. *J. Rubb. Res. Inst. Malaya*, **23**, 47–55.

Nelder, J. A. (1956) Notes on the planning and execution of field experiments. *N.A.A.S. Quart. Rev.*, **34**, 139–46.

Nelder, J. A. (1962) New kinds of systematic designs for spacing experiments. *Biometrics*, **18**, 283–307.

Nelder, J. A. (1966) Evolutionary experimentation in horticulture. *J. Sci. Fd Agric.*, **17**, 7–9.

Neumann, D. (1955) Über die Variabilität der Wuchsleistung von Obstgehölzen in der Baumschule. *Züchter*, **25**, 140–53.

Neumann, D. (1958a) Wuchs- und Ertragserhebungen im Obstbau. *Arch. Gartenb.*, **6**, 3–11.

Neumann, D. (1958b) Beiträge zur Statistik der Wuchs- und Ertragsleistung von Obstbäumen. *Arch. Gartenb.*, **6**, 23–87.

Neumann, U. (1960) Feldversuche mit schwarzen Johannisbeeren. *Arch. Gartenb.*, **8**, 538–58.

Norton, H. W. (1939) The 7×7 squares. *Ann. Eugen.*, **9**, 269–307.

Ollagnier, M. (1951) Forme, dimension des parcelles et nombre de répétitions dans les essais culturaux sur arachide et sur palmier à huile. *Oléagineux*, **6**, 707–10.

Outhwaite, A. D. and Rutherford, A. A. (1955) Covariance analysis as an alternative to stratification in the control of gradients. *Biometrics*, **11**, 431–40.

Paardekooper, E. C. (1966) Results of four uniformity trials to determine the influence of plot size and shape on the efficiency of field experiments on rubber. *Docum. Res. Arch. Rubb. Res. Inst. Malaya* **56**.

Páez, B. G. (1964) Estudios sobre el tamaño y forma de parcelas para ensayos en café. *Proc. Carib. Reg. Amer. Soc. hort. Sci.*, **7**, 69–79.

Páez, B. G. and Siller, L. (1963) The size of experimental cacao plots. *Cacao, Turrialba*, **8**, 11–12.

Papadakis, J. (1937) Méthode statistique pour des expériences sur champ. *Bull. sci. Inst. Amél. Plantes, Salonique*, **23**.

Parker, E. R. (1942) Adjustment of yields in an experiment with orange trees. *Proc. Amer. Soc. hort. Sci.*, **41**, 23–33.

Parker, E. R. and Batchelor, L. D. (1932) Variation in the yields of fruit trees in relation to the planning of future experiments. *Hilgardia*, **7**, 81–161.

Patterson, H. D. (1950) The analysis of change-over trials. *J. agric. Sci.*, **40**, 375–9.

Patterson, H. D. (1951) Change-over trials. *J. roy. statist. Soc.*, **B13**, 256–71.

Patterson, H. D. (1952) The construction of balanced designs for experiments involving sequences of treatments. *Biometrika*, **39**, 32–48.

Pearce, S. C. (1947) The measurement of fruit crops by sampling. *A.R. E. Malling Res. Stat. 1946*, 77–82.

Pearce, S. C. (1948) Randomized blocks with interchanged and substituted plots. *J. roy. statist. Soc.*, **B10**, 252–6.

Pearce, S. C. (1949) The variability of apple trees. I. The extent of crop variation and its minimization by statistical means. *J. hort. Sci.*, **25**, 3–9.

Pearce, S. C. (1952a) Studies in the measurement of apple trees. I. The use of trunk girth to estimate tree size. *A.R. E. Malling Res. Stat. 1951*, 101–4.

Pearce, S. C. (1952b) The design of calibration trials with three varieties. *A.R. E. Malling Res. Stat. 1951*, 105–7.

PEARCE, S. C. (1952c) Some new designs of Latin square type. *J. roy statist. Soc.*, **B14**, 101–6.
PEARCE, S. C. (1955) Some considerations in deciding plot size in field trials with trees and bushes. *J. Indian Soc. agric. Statist.*, **7**, 23–6.
PEARCE, S. C. (1956) Studies in the measurement of apple trees. III. The use of branch girths as a substitute for the trunk girth. *A.R. E. Malling Res. Stat. 1955*, 76–9.
PEARCE, S. C. (1957) Experimenting with organisms as blocks. *Biometrika*, **44**, 141–9.
PEARCE, S. C. (1959) Some recent applications of multivariate analysis to data from fruit trees. *A.R. E. Malling Res. Stat. 1958*, 73–6.
PEARCE, S. C. (1960a) A method of studying manner of growth. *Biometrics*, **16**, 1–6.
PEARCE, S. C. (1960b) Supplemented balance. *Biometrika*, **47**, 263–71.
PEARCE, S. C. (1963) The use and classification of non-orthogonal designs (with discussion). *J. roy. statist. Soc.*, **A126**, 353–77.
PEARCE, S. C. (1964) Experimenting with blocks of natural size. *Biometrics*, **20**, 699–706.
PEARCE, S. C. (1965a) *Biological statistics: an introduction.* McGraw-Hill Book Co., New York.
PEARCE, S. C. (1965b) Statistical techniques in fruit tree research. *Biométrie-Praximétrie*, **6**, 79–92.
PEARCE, S. C. (1965c) The measurement of a living organism. *Biométrie-Praximétrie*, **6**, 143–52.
PEARCE, S. C. (1968) Alternatives to a Latin square. *Biometrics*, **24**, 657–62.
PEARCE, S. C. (1969) Multivariate techniques of use in biological research. *Exp. Agric.*, **5**, 67–77 (reprinted in present book as Appendix II).
PEARCE, S. C. (1970) A comment on totally confounded designs. *Biometrics*, **26**, 329–32.
PEARCE, S. C. (1971) Precision in block experiments. *Biometrika*, **58**, 161–7.
PEARCE, S. C. (1975) Row-and-column designs. *Appl. Statist.*, **24**, 60–74.
PEARCE, S. C. and BROWN, A. H. F. (1960) Improving fruit tree experiments by a preliminary study of the trees. *J. hort. Sci.*, **35**, 56–65.
PEARCE, S. C., CALIŃKSI, T. and MARSHALL, T. F. DE C. (1974) The basic contrasts of an experimental design with special reference to the analysis of data. *Biometrika*, **61**, 449–60.
PEARCE, S. C. and DAVIES, G. E. (1954) Studies in the measurement of apple trees. II. The height at which trunk girths should be measured. *A.R. E. Malling Res. Stat. 1953*, 138–40.
PEARCE, S. C. and DOBERŠEK-URBANC, S. (1967) The measurement of irregularity in growth and cropping. *J. hort. Sci.*, **42**, 295–305.
PEARCE, S. C. and HOBLYN, T. N. (1948) A review of experimental design at East Malling. *A.R. E. Malling Res. Stat. 1947*, 88–100.
PEARCE, S. C. and HOLLAND, D. A. (1960) Some applications of multivariate methods in botany. *Appl. Statist.*, **9**, 1–7.
PEARCE, S. C. and HOLLAND, D. A. (1961) Analyse des composantes, outil en recherche biométrique. *Biométrie-Praximétrie*, **2**, 159–77.
PEARCE, S. C. and JEFFERS, J. R. N. (1971) Block designs and missing data. *J. roy. statist. Soc.*, **B33**, 131–6.
PEARCE, S. C. and MOORE, C. S. (1962) A study of sources of variation in apple tree experiments. *J. hort. Sci.*, **37**, 175–89.
PEARCE, S. C. and TAYLOR, J. (1948) The changing of treatments in a long-term trial. *J. agric. Sci.*, **38**, 402–10.
PEARCE, S. C. and TAYLOR, J. (1950) The purpose and design of calibration trials. *A.R. E. Malling Res. Stat. 1949*, 83–90.
PEARCE, S. C. and THOM, J. M. S. (1951a) The variability of apple trees. II. The optimum size for unguarded plots. *J. hort. Sci.*, **26**, 98–108.
PEARCE, S. C. and THOM, J. M. S. (1951b) A study of plot size with Nigerian estate cacao. *J. hort. Sci.*, **26**, 261–7.
PEARSON, E. S. and HARTLEY, H. O. (1951) Charts of the power function of the analysis of variance tests derived from the non-central *F*-distribution. *Biometrika*, **38**, 112–30.
PIERIS, W. V. D. and SALGADO, M. L. M. (1937) Experimental error in field experiments with coconuts. *Trop. Agric., Trin.*, **89**, 75–85.
PIMENTEL-GOMES, F. (1951a) A interpolação de Lei de Mitscherlich e a análise da variância em experiências de adubação. *An. Esc. sup. Agric. Queiroz*, **8**, 185–204.
PIMENTEL-GOMES, F. (1951b) A Lei de Mitscherlich e a análise da variância em experiências de adubação. *An. Esc. sup. Agric. Queiroz*, **8**, 355–68.
PIMENTEL-GOMES, F. (1953) The use of Mitscherlich's regression law in the analysis of experiments with fertilizers. *Biometrics*, **9**, 498–516.
PIMENTEL-GOMES, F. (1963) Diseño experimental en frutales. *Rev. Agric., Piracicaba*, **38**, 207–24.
POUND, F. J. (1932) The fruitfulness of cacao. The significance of the relation between total variance and mean annual yield. *Trop. Agric., Trin.*, **9**, 288–90.
PREECE, D. A. (1971) Iterative procedures for missing values in experiments. *Technometrics*, **13**, 743–53.
PREECE, D. A. (1972) Non-additivity in two-way classifications with missing values. *Biometrics*, **28**, 574–7.

PREECE, D. A. and GOWER, J. C. (1974) An iterative computer procedure for mixed-up values in experiments. *Appl. Statist.*, **23**, 73–4.

PROEBSTING, E. L. (1942) The relative yields of border fruit trees. *Proc. Amer. Soc. hort. Sci.*, **41**, 34–6.

QUENOUILLE, M. H. (1948) The analysis of covariance and non-orthogonal data. *Biometrics*, **4**, 240–6.

QUENOUILLE, M. H. (1953) *The design and analysis of experiment*. Charles Griffin & Co. Ltd., London.

REED, H. S. (1929) Correlations between growth and fruit production of apricots. *Proc. Amer. Soc. hort. Sci.*, **25**, 247–9.

RICH, S. (1952) Using half-tree plots for increasing the efficiency of fungicide tests. *Phytopathology*, **42**, 353–4.

RIGNEY, J. A. (1946) Some statistical problems confronting horticultural investigators. *Proc. Amer. Soc. hort. Sci.*, **48**, 351–7.

RIGNEY, J. A., MORROW, E. B. and LOTT, W. L. (1949) A method of controlling experimental error for perennial horticultural crops. *Proc. Amer. Soc. hort. Sci.*, **54**, 209–12.

RIVES, M. (1960) Sur l'analyse de la variance. I. Emploi de transformations. *Ann. Amél. Plantes*, **10**, 309–31.

RIVES, M. (1962) Sur l'analyse de la variance. II. Additivité. *Ann. Amél. Plantes*, **12**, 5–18.

RIVES, M. (1963) Description de l'hétérogénéité des champs d'expérience. *Biom. Z.*, **11**, 113–22.

ROACH, W. A. (1938) Plant injection for diagnostic and curative purposes. *Tech. Commun. imp. Bur. Hort., E. Malling*, **10**.

ROACH, W. A. and ROBERTS, W. O. (1945) Further work on plant injection for diagnostic and curative purposes. *Tech. Commun. imp. Bur. Hort., E. Malling*, **16**.

ROESSLER, E. B. (1943) Valid estimates of variance in the analysis of pooled data. *Proc. Amer. Soc. hort. Sci.*, **42**, 481–3.

ROESSLER, E. B. and LEACH, L. D. (1944) Analysis of combined data for identical replicated experiments. *Proc. Amer. Soc. hort. Sci.*, **44**, 323–8.

ROGERS, H. H. and KNIGHT, R. (1953) Some observations on yield trials of cacao at Tafo. *Proc. W. Afr. Inst. Cacao Res. Conf. Tafo*, 42–5.

ROGERS, I. S. (1972) Practical considerations in the use of systematic spacing designs. *Aust. J. exp. Agric. Anim. Husb.*, **12**, 306–9.

ROGERS, W. S. (1931) Recording apparatus for horticultural experiments including automatic counting devices. *A.R. E. Malling Res. Stat. 1928, 1929 and 1930. II Suppl.*, 65–73.

ROGERS, W. S. and BEAKBANE, A. B. (1957) Stock and scion relations. *Ann. Rev. Plant Phys.*, **8**, 217–36.

ROGERS, W. S. and EDGAR, J. L. (1938) Strawberry cultivation studies. I. The performance of individual plants of clonal families. *J. Pomol.*, **16**, 63–90.

ROGERS, W. S. and VYVYAN, M. C. (1928) The root systems of some ten-year-old apple trees on two different root stocks, and their relation to tree performance. *A.R. E. Malling Res. Stat. 1926–1927. II Suppl.*, 31–43.

ROGERS, W. S. and VYVYAN, M. C. (1934) Root studies. V. Rootstock and soil effect on apple root systems. *J. Pomol.*, **12**, 110–50.

ROOSJE, G. S. (1963) Overwegingen bij het aanleggen van proefvelden en hetopzetten van veldproeven in de fruitteelt. *Meded. Dir. Tuinb.*, **26**, 136–50.

RUBIN, D. R. (1972) A non-iterative algorithm for least squares estimation of missing values in any analysis of variance design. *Appl. Statist.*, **21**, 136–41.

SALGADO, M. L. M. (1946) Recent studies on the manuring of coconuts in Ceylon. *Trop. Agriculturist*, **102**, 149–54.

SATTERTHWAITE, F. E. (1946) An approximate distribution of estimates of variance components. *Biometrics Bull.*, **2**, 110–4.

SCHEFFÉ, H. (1959) *The analysis of variance*. John Wiley & Sons, New York.

SCHULTZ, E. F., JR. and SCHNEIDER, G. W. (1955) Sample size necessary to estimate size and quality of fruit, growth of trees and per cent fruit set of apples and peaches. *Proc. Amer. Soc. hort. Sci.*, **66**, 36–44.

SEN, A. R. (1963a) Use of pretreatment data in designing experiments on tea. *Emp. J. exp. Agric.*, **31**, 41–9.

SEN, A. R. (1963b) Some techniques of experimentation with clonal tea based on a uniformity trial. *Emp. J. exp. Agric.*, **31**, 296–310.

SEN, A. R. and BISWAS, A. K. (1966) Some techniques of experimentation with tea in north-east India. *Exp. Agric.*, **2**, 89–100.

SHAH, S. M. I. (1950) Influence of the number of trees per plot on the precision of peach yield trials. *Punjab Fruit J.*, **14**, 17–22.

SHARPE, R. H. and BLACKMON, G. H. (1950) A study of plot size and experimental design with pecan yield data. *Proc. Amer. Soc. hort. Sci.*, **56**, 236–41.

SHARPE, R. H. and WINSOR, H. W. (1951) Cross-feeding and boron placement studies with pecan. *Proc. Amer. Soc. hort. Sci.*, **57**, 202–6.

SHAW, J. K. (1942) Trunk diameters of young apple trees on clonal stocks. *Proc. Amer. Soc. hort. Sci.*, **40**, 269–71.

SHRIKANDE, V. J. (1958) Some considerations in designing experiments on coconut trees. *Indian Coconut J.*, **11**, 140–56.
SILVEY, R. J. E. (1944) Methods of listener research employed by the British Broadcasting Corporation. *J. roy. statist. Soc.*, **107**, 190–230.
SIMPSON, R. C. (1936) Individual tree yield records in a commercial apple orchard. *Proc. Amer. Soc. hort. Sci.*, **34**, 155–9.
SMITH, H. F. (1938) An empirical law describing heterogeneity in the yields of agricultural crops. *J. agric. Sci.*, **28**, 1–23.
SMITH, H. F. (1950) Error variance of treatment contrasts in an experiment with missing observations (with special reference to incomplete Latin squares). *J. Indian Soc. agric. Statist.*, **2**, 111–24.
SNEDECOR, G. W. (1956) *Statistical methods.* Collegiate Press, Menasha, Wisc., U.S.A.
SPRENT, P. (1970) Some problems of statistical consultancy (with discussion). *J. roy. statist. Soc.*, **A133**, 139–64.
STEEL, R. G. D. (1955) An analysis of perennial crop data. *Biometrics*, **11**, 201–12.
STEVENS, W. L. (1949) Análise estatistíca do ensaio de variedades de café. *Bragantia*, **9**, 103–23.
STEVENS, W. L. and FRAGA, C. A. (1951) Análise do ensaio de desbaste de ramos inferiores do cafeeiro. *Bragantia*, **11**, 285–9.
STEWART, A. B. (1947) Planning and conduct of fertilizer experiments. *Colonial* **214** (H.M.S.O., London), 3–9.
STRICKLAND, A. G. (1935) Error in horticultural experiments. *J. Dep. Agric. Vict.*, **33**, 408–16.
STRICKLAND, A. G., FORSTER, H. C. and VASEY, A. J. (1932) A vine uniformity trial. *J. Dep. Agric. Vict.*, **30**, 584–93.
SUDDS, R. H. and ANTHONY, R. D. (1929) The correlation of trunk measurements with tree performance in apples. *Proc. Amer. Soc. hort. Sci.*, **25**, 244–6.
SUTTON, B. G. and POTTER, G. F. (1954) In *Mineral nutrition of fruit crops* (Ed. N. F. Childers), 569–83. Rutgers University, N.J., U.S.A.
SZCZEPASŃKI, K. and KŁOSSOWSKI, W. (1963) Ocena wysokości plonu jabłoni w doświadczeniach polowych. *Prace Inst. Sadown. w Skierniewicach*, **7**, 261–76.
TAYLOR, R. H. (1926) The inter-relationship of yield and the various vegetative characters in *Hevea brasiliensis*. *Bull. Dep. Agric. Ceylon*, **77**.
TAYLOR, J. (1948) Errors of treatment comparisons when observations are missing. *Nature* (*Lond.*), **162**, 262.
TAYLOR, J. (1949) A valid restriction of randomization for certain field experiments. *J. agric. Sci.*, **39**, 303–8.
TAYLOR, J. (1950) The comparison of pairs of treatments in split-plot experiments. *Biometrika*, **37**, 443–4.
TAYLOR, J. (1951a) The estimation of fruit size of cherries by sampling methods. *A.R. E. Malling Res. Stat. 1950*, 93–9.
TAYLOR, J. (1951b) Statistical studies on strawberry crop and vigour measurements. *A.R. E. Malling Res. Stat. 1950*, 100–7.
TIDBURY, G. E. (1943) Stem girth as a criterion in field trials with young clove trees. *Emp. J. exp. Agric.*, **11**, 33–7.
TIKU, M. L. (1967) Tables of the power of the *F*-test. *J. Amer. statist. Ass.*, **62**, 525–39.
TIKU, M. L. (1972) More tables of the power of the *F*-test. *J. Amer. statist. Ass.*, **67**, 709–10.
TIPPETT, L. H. C. (1931) *The methods of statistics.* Williams and Norgate Ltd., London.
TRAIL, S. M. and WEEKS, D. L. (1973) Extended complete block designs generated by BIBD. *Biometrics*, **29**, 565–78.
TUKEY, J. W. (1949a) Comparing individual treatment means in the analysis of variance. *Biometrics*, **5**, 99–114.
TUKEY, J. W. (1949b) One degree of freedom for non-additivity. *Biometrics*, **5**, 232–42.
TUKEY, J. W. (1963) The future of data analysis. *Ann. math. Statist.*, **33**, 1–67.
TURRELL, F. M. (1946) Estimating heights of citrus trees. *Proc. Amer. Soc. hort. Sci.*, **48**, 147–50.
TURRELL, F. M., GARBER, M. J., JAMES, W. W., COOPER, W. C. and YOUNG, R. H. (1969) Growth equations and curves for citrus trees. *Hilgardia*, **39**, 429–45.
UPSHALL, W. H. (1939) Transplanting stock in peach seedling rootstocks and its effect on size of nursery and orchard trees. *Proc. Amer. Soc. hort. Sci.*, **37**, 340–2.
URUI, K. and LILLEHARD, O. (1959) Limb circumference measurements as an aid in the better evaluation of apple fruit thinning data. *Proc. Amer. Soc. hort. Sci.*, **74**, 67–71.
VAIDYANATHAN, M. (1934) The method of "covariance" applicable to the utilization of the previous crop records for judging the improved precision of experiments. *Indian J. agric. Sci.*, **4**, 327–41.
VAN DER KNAAP, W. P. (1955) Blancoproeven met cacao en koffie. *Arch. Koffiecult.*, **17**, 187–239.
VAN DER PLANCK (1963) *Plant diseases: epidemics and control.* Academic Press, New York and London.
VERNON, A. J. (1968) Edge effects in a cocoa shade and manurial experiment. *J. hort. Sci.*, **43**, 147–52.
VERNON, A. J. (1969) The description of yield curves of multi-harvest crops. *J. hort. Sci.*, **44**, 13–25.

VERNON, A. J. and MORRIS, M. G. (1964) Variability in West African peasant cocoa farms; time studies on Agodi cocoa yields. *J. hort. Sci.*, **39**, 241–53.

VISSER, T. (1958) The validity of assessing tea yields on a basis of intermittent plucking and test plucking. *Tea Q.*, **29**, 21–9.

VON NEUMANN, T. (1941) Distribution of the ratio of the mean square successive difference to the variance. *Ann. math. Statist.*, **12**, 367–95.

WADLEIGH, C. H. and THARP, W. H. (1940) Factorial design in plant nutrition experiments in the greenhouse. *Bull. Ark. agric. Exp. Stat.* **401**.

WALLOFF, N. (1966) Scotch broom (*Sarothamnus scoparius* (L.) Wimmer) and its insect fauna introduced into the Pacific Northwest of America. *J. appl. Ecol.*, **3**, 293–311.

WARD, G. C. and DICK, I. D. (1952) Non-additivity in randomized block designs and BIB designs. *N.Z. J. Sci. Techn.*, **33**, 430–5.

WEBBER, H. J. (1932) Variation in citrus seedlings and their relation to rootstock selection. *Hilgardia*, **7**, 1–79.

WEBSTER, C. C. (1939) A note on a uniformity trial with oil palms. *Trop. Agric., Trin.*, **16**, 15–19.

WELCH, B. L. (1947) A generalization of "Student's" problem when several different population variances are involved. *Biometrika*, **34**, 28–35.

WELCH, B. L. (1949) Further note on Mrs. Aspin's tables and on certain approximations to the tabled function. *Biometrika*, **36**, 293–6.

WELLENSIEK, S. J. (1938) Onderzoekingen over quantitatieve theeselectie. IV. Moederboomselectie. *Arch. Theecult.*, **12**, 2–70.

WELLMAN, R. H., THURSTON, H. W., JR. and WHALEY, F. H. (1948) A method for correcting the geographic variation in field experiments. *Contr. Boyce Thompson Inst.*, **15**, 153–64.

WESTGARTH, D. R. (1953) Methods of improving precision in hevea trials. *Arch. Rubbercult.*, Extra No., May 1953, 180–93.

WESTWOOD, M. N. and ROBERTS, A. N. (1970) The relationship between trunk cross-sectional area and weight of apple trees. *Proc. Amer. Soc. hort. Sci.*, **95**, 28–30.

WILCOX, A. N. (1927) A study of field plot technic with strawberries. *Sci. Agric.*, **8**, 171–4.

WILCOX, J. C. (1937a) Field studies of apple tree growth and fruiting. I. Sampling and measuring terminal shoots. *Sci. Agric.*, **17**, 563–72.

WILCOX, J. C. (1937b) Field studies of apple tree growth and fruiting. II. Correlations between growth and fruiting. *Sci. Agric.*, **17**, 573–86.

WILCOX, J. C. (1937c) Field studies of apple tree growth and fruiting. III. Some observations on the measurement of tree vigour. *Sci. Agric.*, **17**, 657–69.

WILCOX, J. C. (1940) Adjusting apple yields for differences in size of tree. *Sci. Agric.*, **21**, 139–48.

WILCOX, J. C. (1945) Some factors affecting apple yields in the Okanagan Valley. I. Tree size, tree vigour, biennial bearing, and distance of planting. *Sci. Agric.*, **25**, 189–213.

WILLIAMS, C. B. (1937) The use of logarithms in the interpretation of certain entomological problems. *Ann. appl. Biol.*, **24**, 404–14.

WILLIAMS, E. J. (1952) The interpretation of interactions in factorial experiments. *Biometrika*, **39**, 65–81.

WISHART, J. (1936) Tests of significance in analysis of covariance. *J. roy. statist. Soc., Suppl.* **3**, 79–82.

WISHART, J. (1938) Field experiments of factorial design. *J. agric. Sci.*, **28**, 299–306.

WISHART, J. (1950) Field trials II. The analysis of covariance. *Tech. Commun. Commonw. Bur. Plant Breed. Genet., Cambridge*, **15**.

WISHART, J. and SANDERS, H. G. (1936) *Principles and practice of field experimentation.* Emp. Cotton Grow. Corp., London. Reissued as *Tech. Commun. Commonw. Bur. Plant Breeding, Cambridge*, **18** (1955).

WORTHINGTON, B. A. (1975) General iterative method for analysis of variance when block structure is orthogonal. *Biometrika*, **62**, 113–19.

YARRANTON, G. A. (1967) Principal component analysis of data from saxicolous bryophyte vegetation at Steps Bridge, Devon. *Canad. J. Bot.*, **45**, 93–115.

YATES, F. (1933) The analysis of replicated experiments when the field results are incomplete. *Emp. J. exp. Agric.*, **1**, 129–42.

YATES, F. (1935) Complex experiments (with discussion). *J. roy. statist. Soc., Suppl.* **2**, 181–247.

YATES, F. (1936a) Incomplete Latin squares. *J. agric. Sci.*, **26**, 301–15.

YATES, F. (1936b) Incomplete randomized blocks. *Ann. Eugen.*, **7**, 121–40.

YATES, F. (1937) The design and analysis of factorial experiments. *Tech. Commun. Commonw. Bur. Soil Sci., Rothamsted*, **35**.

YATES, F. (1938) The recovery of inter-block information in balanced incomplete block designs. *Ann. Eugen.*, **10**, 317–25.

YATES, F. (1940) Modern experimental design and its function in plant selection. *Emp. J. exp. Agric.*, **8**, 223–30.

YATES, F. (1946) The place of statistics in agricultural research. *Agric. Progr.*, **21**. (Seen only in reprint form. Page numbers unknown.)

YATES, F. and WATSON, D. J. (1934) Observer's bias in sampling observations in wheat. *Emp. J. exp. Agric.*, **2**, 174–7.

YEAGER, A. F. and LATIMER, L. P. (1940) Tree girth and yield as indicators of subsequent apple yield productivity. *Proc. Amer. Soc. hort. Sci.*, **37**, 101–5.

YOUDEN, W. J. (1937) Use of incomplete block replications in estimating tobacco-mosaic virus. *Contr. Boyce Thompson Inst.*, **9**, 41–8.

YOUNGE, O. R. and PLUNCKNETT, D. L. (1969) Estimating the number of papaw (*Carica papaya*) trees required for reliable yield data. *Qd J. agric. Anim. Sci.*, **26**, 289–92.

YULE, G. U. (1927) On reading a scale. *J. roy. statist. Soc.*, **90**, 570–87.

INDEX

	Sections
Acceptability of fruit	97
Accuracy	*see* Standard error
Adaptability of design	43
Adjustment by neighbouring plots	45
Adjustment of treatment means	30, 31, 41, 45, 73, App. III
Algorithms	70, App. III
Analysis of data	
Analysis of covariance	35, 41, 45, 80, 81, 82, 83, 93, App. IV
Analysis of variance	11, 21, 22, 24, 25, 26, 27, 30, 33, 74, 83, 84, 86, App. I, App. III
Correlation analysis	App. IV
Multivariate analysis	74, App. II
Regression analysis	App. IV
Angles of equal information	73
Apples	14, 15, 43, 52, 60
Avocado pears	52
Back-transformation	73
Biennial effects	43, 74
Blocks	
Dispersed	13, 65, 75
Extended	30
Incomplete	21, 30
Plants as	35
Purpose of	20, 41, 50
Size and shape	50
Blossom records	96
Blueberries	43
Blunders	80
Buffers	*see* Guards, etc.
Cacao	*see* Cocoa
Calibration	40, 41, 43, 44, 50
Categories	90, 93
Cherries	14, 50, 51
Citrus	14, 15, 43, 52
Cloves	43
Cocoa	14, 15, 43, 52, 53, 82
Coconuts	41, 43, 52, 60
Coffee	43, 51, 52, 74
Computations	70, 83, App. I
Concurrences	30
Confounding	36
Contrasts	20, 24, 26, 36, 73, 74
Crop records	91
Damaged experiments	80
Degrees of freedom	11, 20, 24, 30, 73, 83
Designs	
Balanced incomplete blocks	31, 35
Calibration	43
Change-over	34
Completely randomized	20
Complexity of	38
Designs (*cont.*)	
Criss-cross	26, 61
Definitive	60
Disconnected	86
Fan	66
Flexibility of	21, 22, 23
Group-divisible	32, 33
with Interchanged plots	32, 62
Latin squares	22, 23, 33
Multiple Latin squares	22
Non-orthogonality of	30, 84
Nonce	39
Orthogonality of	21, 22, 30
Randomized blocks	20, 21
Robustness of	21, 22, 80
Row-and-column	33
Semi-Latin squares	39
Split-plot	25, 61, 72
Stripe	42
with Substituted plots	32, 62
with Supplemented balance	32, 33, 39, 62, App. III
Tentative	60
Tile	42
Totally balanced	31, 33
Youden squares	33
Discontinuity corrections	73
Effects (*see also* Interactions)	
Local	35
Main	24, 26
Remote	35
Efficiency	30, 31, 32, 36, App. III
Error, Heterogeneity of	*see* Heterogeneity of error
Error, Standard	*see* Standard error
Examination of data	70
Experiments	
Commercial fields used for	13, 41, 44
of Commercial systems	65
Damaged	80
Definitive	60
on Disease spread	67
Dusting	62
Error of	11, 14, 26, 31, 40, 70
Factorial	*see* Factors
Fertilizer	60
Herbicide	61
Interpretation of	20, 70
Irrigation	61
Long-term	14, 40
Objects of	11, 25
Preliminary	54
Pruning	64
Range of	13
Rootstock	63
Soil management	61
Spacing	66

	Sections
Experiments (*cont.*)	
Spraying	62
Tentative	60, 62
Variety	63
Factors	24, 25, 33, 60, 67, 72
Flexibility of design	21, 22, 23
Frameworking	63
Grapes	14, 43, 50, 51, 52, 60
Guard rows, etc.	30, 51, 52, 53, 60, 61, 62
Heterogeneity	
of Environmental conditions	14, 50, 51, 52
of Error	73, 82
Heteroscedasity	*see* Heterogeneity of error
Homoscedasity	*see* Heterogeneity of error
Hops	52, 82
Injection techniques	60
Interactions	24, 25, 26, 27, 30, 36, 70, 71, App. IV
Iteration	83, App. III
Interblock information	20, 35
Least significant difference	72
Lemons	*see* Citrus
Logarithmic transformation	73
Main effects	*see* Effects, main
Missing plots	30, 51, 83, 84
Mixed model	24
Mixed-up plots	85
Multiple comparisons	71, 72
Multivariate methods	*see* Analysis of data
Nomination	71
Oil palms	43, 52, 74
Oranges	*see* Citrus
Orthogonality	*see* Designs, Orthogonality of
Papadakis, Method of	45
Papaw	52
Partition of	
Error sum of squares	27, 73
Latin square	23
Total sum of squares	*see* Analysis of data
Treatment sum of squares	24, 30, 70, 86, App. IV
Passion fruit	53
Peaches	14, 15, 52
Pears	14, 43, 52
Pecans	15, 43, 52, 60
Pepper	52
Perennial plants, Statistical problems of	14, 71
Photographic records	91
Pineapples	52
Plants, Selection of	*see* Selection
Plots	
Damaged	82
Interchanged	86
Missing	30, 83, 84
Mixed-up	85
Nature of	11, 35, 40, 44, 51
Parts of tree as	35
Substituted	32
Size and shape	43, 51, 65, 91
Plums	51
Power of tests	12
Precision	*see* Standard error
Pseudo-variates	81, 83
Randomization	11, 14, 15, 16, 20, 21, 23, 25, 26, 27, 34, 60, 61, 64
Recording	90, 91, 98
Rectangularity	66
Replants	82
Replication	
Actual	22
Degree of	54, 55, 80
Effective	31, 32
Hidden	54
Response surfaces	24, 60
Residual effects	33, 34
Robustness of design	*see* Design
Rubber	43, 51, 52, 53
Sampling	90, 92
Selection	
of plants	14, 15, 40, 63
of sites	14, 75
Significance	11, 12, 70, 71
Significant difference	72
Size records	41, 43, 95
Square-root transformation	73
Standard error	11, 30, 31, 32, 40, 54, 70, 72, 84
Statistics in agricultural science	10, 11, 16
Strawberries	15, 43, 52, 82
Sub-plots, etc.	25, 53
Surveys	11, 13
Tea	43, 50, 51, 52
Transformations	41, 73, 74
Tying of Latin squares	22
Tests	
F-test	11, 20, 21, 24, 72, 73
Multiple range	*see* Multiple comparisons
Power of	12
t-test	72, 73
Treatments	
Choice of	13, 16
Changing of	23, 40, 60
Equivalence of	31

	Sections
Treatments (*cont.*)	
Factorial	*see* Factors
Means of	*see* Adjustment of treatment means
Orthogonality of, to blocks	30
Proportionality of	30, 86
Virtual identity of	72
Trials	*see* Experiments
Variance, Analysis of	*see* Analysis of data
Variation	*see* Experiments, Error of
Walnuts	52

SOME OTHER PUBLICATIONS OF THE COMMONWEALTH BUREAU OF HORTICULTURE AND PLANTATION CROPS

JOURNALS

Horticultural Abstracts. Issued quarterly since 1931, monthly from 1973. Annual subscription £78·00 ($185.00) to subscribers in non-contributing countries and £35·00 to subscribers in contributing countries. Composite indexes up to 1965 are also available.

Ornamental Horticulture. Issued monthly from 1975. Annual subscription £15·00 ($36.00).

TECHNICAL COMMUNICATIONS

T.C.33. Mist propagation of cuttings, by J. E. K. Cox, 1974, £4·00 ($10.40).

T.C.34. An annotated bibliography of colour illustrated mineral deficiency symptoms in tropical crops, by J. B. D. Robinson 1974, £5·00 ($13.00).

T.C.35. French–English Horticultural Dictionary, by D. O'D. Bourke, 1974, £5·00 ($13.00).

RESEARCH REVIEWS

R.R.2. Crop responses to water at different stages of growth, by P. J. Salter and J. E. Goode, 1967, £2·25 ($5.90).

R.R.3. The biology of apple and pear storage, by J. C. Fidler, B. G. Wilkinson, K. L. Edney, and R. O. Sharples, 1973, £5·00 ($13.00).

R.R.4. Root temperature and plant growth, by A. J. Cooper, 1973, £1·60 ($4.20).

R.R.5. Climate and the orchard, edited by H. C. Pereira, 1975, £5·00 ($13.00).

HORTICULTURAL REVIEWS

H.R.3. Deciduous fruit tree cultivars for tropical and sub-tropical regions, by H. C. Ruck, 1975, £2·75 ($6.90).

C.A.B. JOURNALS AND OTHER PUBLICATIONS

The Commonwealth Agricultural Bureaux provide a comprehensive coverage of the world's literature on all aspects of agricultural science in a series of abstract journals devoted to: AGRICULTURAL ECONOMICS, AGRICULTURAL PARASITOLOGY, ANIMAL BREEDING AND GENETICS, ANIMAL HEALTH, NUTRITION, DAIRY SCIENCE AND TECHNOLOGY, ENTOMOLOGY, FIELD CROPS, FORESTRY, HORTICULTURE AND PLANTATION CROPS, MYCOLOGY, PASTURES, PLANT BREEDING AND GENETICS, SOILS AND FERTILIZERS.

Full details of these journals and of numerous occasional publications, in which research topics of current interest are reviewed, can be obtained from:

COMMONWEALTH AGRICULTURAL BUREAUX,
CENTRAL SALES BRANCH, FARNHAM HOUSE,
FARNHAM ROYAL, SLOUGH SL2 3BN, ENGLAND